Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER

DAVID A. GRANT, Associate Editor
UNIVERSITY OF WISCONSIN

DELOS D. WICKENS, Associate Editor
OHIO STATE UNIVERSITY

CONSULTING EDITORS

E. James Archer, *University of Wisconsin*
Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Paul M. Fitts, *Ohio State University*
Frederick C. Frick, *Massachusetts Institute of Technology*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
Harold W. Hake, *University of Illinois*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Howard H. Kendler, *New York University*
Donald B. Lindsley, *University of California, Los Angeles*
Kenneth MacCorquodale, *University of Minnesota*
Quinn McNemar, *Stanford University*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*

Arthur C. Hoffman, *Managing Editor*
Helen Orr, *Assistant Managing Editor*
Frances Hasel and Sadie J. Doyle, *Editorial Staff*

CONTENTS

Probability and Work as Determiners of Multichoice Behavior: M. H. DETAMBEL AND L. M. STOLUROW 73

Relations of Creative Responses to Working Time and Instructions: P. R. CHRISTENSEN, J. P. GUILFORD, AND R. C. WILSON 82

Effects of Experience on Perception of Causality: H. E. GRUBER, C. D. FINK, AND V. DAMM 89

A Measure of Stimulus Similarity and Errors in Some Paired-Associate Learning Tasks: E. Z. ROTHKOPF 94

The Role of Context Stimuli in Verbal Learning: L. R. PETERSON AND M. J. PETERSON 102

Strength of Secondary Reinforcement as a Determiner of the Effects of Duration of Goal Response on Learning: D. R. POWELL, JR., AND C. C. PERKINS, JR. 106

Muscular Tension and the Human Blink Rate: D. C. KING AND K. M. MICHELS 113

Effect of Food, Drink, and Tobacco Deprivation on the Conditioning of the Eyeblink Response: C. M. FRANKS 117

The Effect on Problem Solving of Success or Failure as a Function of Cue Specificity: R. J. RHINE 121

Motivational Properties of Frustration: III. Relation of Frustration Effect to Antedating Goal Factors: A. AMSEL AND W. HANCOCK 126

Anchor Effects in Absolute Judgments: C. W. ERIKSEN AND H. W. HAKE 132

Stimulus Generalization After Equal Training on Two Stimuli: H. I. KALISH AND N. GUTTMAN 139

Dynamic Contour Perception: W. M. SMITH AND W. L. GULICK 145

Erratum 152

American Psychological Association

Vol. 53 No. 2 February 1957

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal, experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research effort is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, 53, 1–2).

Address all articles submitted for publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, P. O. Box 1557, Lackland Air Force Base, Texas. All manuscripts should be submitted in two complete copies; only the original copy will be returned if the article is not accepted for publication. Figures may be original line drawings or glossy photographic prints, but should not exceed 8½ by 11 in. Care should be taken to insure legibility of labels and critical features of drawings when reduced to a single-column or (exceptionally) a double-column figure in the JOURNAL.

To save printing costs supplementary material to articles in this JOURNAL are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 53, No. 2 FEBRUARY, 1957

PROBABILITY AND WORK AS DETERMINERS OF MULTICHOICE BEHAVIOR [1]

MARVIN H. DETAMBEL [2] AND LAWRENCE M. STOLUROW [3]

University of Illinois

In trouble shooting, a symptom of malfunction frequently has a probabilistic relationship to several possible troubles each of which requires a different amount of work to check. The task in such a situation is to check the possible troubles, one at a time, until the actual trouble is discovered (2).

This sort of trouble-shooting task was simulated by Detambel (1). He repeatedly confronted *S* with four decks of cards; one deck contained a "trouble card," and each of the others a card representing the absence of trouble. Decks differed both in the probability of their containing the "trouble" and in the number of cards that had to be turned to discover the indicator card. However, since all of the 24 possible sequences for checking the four decks required the same mean number of cards to be turned per discovery of the trouble, they all were equally efficient. In this situation, *Ss* tended increasingly to select the previously unchecked deck which had the greatest probability of containing the trouble, even though that alternative also required the most work. Since all checking sequences were equally efficient, this preference could not be attributed to the relative efficiency of the preferred sequence. Rather, it indicated that work and probability variables were differentially weighted as behavior determiners.

If such a bias persisted when the combinations of work and probability made it more efficient, on the average, to check the low probability—low work alternative, then *Ss* would do more work than necessary. Such a result would be of both theoretical and practical interest.

The two purposes of the present study were to verify the results of the previous study and to determine the effects of trouble-shooting experience in which (*a*) checking the high probability—high work alternative [the *Ss*' preferred strategy in (1)] results in the greatest amount of work per discovery of the trouble, and conversely (*b*) checking the low probability—low work alternative results in the most efficient strategy.

METHOD

A total of 24 paid, male undergraduate students at the University of Illinois served as

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25726, monitored by the Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government. The authors are grateful to Mrs. Sally Beck and Mr. William LeMar for their assistance in this study.

[2] Now at Douglass Aircraft Co., Santa Monica, California.

[3] Now at Propulsion Research Unit, Maintenance Laboratory Field Unit No. 1, Air Force Personnel and Training Research Center, Chanute Air Force Base, Illinois.

Ss. Each *S* sat at a table that held a rectangular device with three fiberboard lids, each fastened to the top surface by one or more screws. The *Ss*' task with respect to this device is described in the following instructions, which were read to every *S* at the beginning of the experimental period:

"What you are about to do is similar in certain respects to the work of a mechanic. We will bring you a series of pieces of equipment like the one you see on the table. There is something wrong with each piece of equipment, and it will be your job to find the trouble and repair it.

"The trouble can occur in any of three sections of the equipment. Sometimes you will find it in this section (point to it), sometimes in this section, and sometimes in this section. You can find out whether the trouble is in a particular section by removing the screws, lifting the cover plate and looking inside. If you see the letters 'OK' you will know that section of the equipment is all right. But if a piece of tape covers the letters, you have found the trouble. For each piece of equipment you are to find the trouble and repair it by removing the tape.

"There will be nearly 200 pieces of equipment for you to repair. After you have repaired about a quarter of them, we'll take a break and then repair some more. For the morning's (afternoon's) work you will receive 4 hr. pay, but you may leave in less than 4 hr. if you have completed all of the repairs.

"As you perform this task, you should notice two things, the amount of work you must do in order to check each section and the number of times the trouble occurs in each section. Paying attention to the amount of work and the frequency of trouble in each section may help you to do the job quickly. We want you to complete this work with the least possible time and effort."

The apparatus to be "repaired" was a rectangular wooden frame with slots for four brass plates; however, only three were used in this experiment. In each plate there was a hole of 2.5 in. diameter and, regularly spaced about the hole's circumference, eight screw holes. At the beginning of every trial each of the three holes was covered by a circular fiberboard lid, $\frac{1}{4}$ in. thick, held in place by one or more $\frac{3}{4}$-in. round head-screws. Screws were inserted flush with the surface of the lid, but not so tightly as to prevent their removal without the aid of a screw driver.

After removing the screws *S* could lift the lid, exposing either the "OK" sign or the tape that represented the trouble. On each trial *S* removed as many lids as necessary to discover and repair the trouble.

Two devices were used so that while one was in use, an assistant could both insert the new trouble in the other and reassemble it for the next trial. As soon as both *S* and the assistant completed their work on a trial, the next one was begun.

For most *Ss* there was a rest period of approximately 12 min. following each block of 48 trials, three such rest periods occurring during the 192 trials. However, three of the 12 *Ss* in each group completed the first two blocks of 48 trials on one day, and the remaining two blocks on another day not more than three days later.

After each break, before resuming the experimental task, *S* was reminded that he was to complete the set of repairs, doing as little work as possible, and that to help him accomplish this he should pay attention to the frequency of trouble and the amount of work in each section of the equipment.

Table 1 shows for both tasks, the trouble probability and the amount of work (number of screws) associated with each section of the equipment. Task 3-2-1 is comparable to the earlier experiment (1) in that all checking sequiences are about equally efficient, i.e., all six checking sequences result in removal of the same mean number of screws per discovery of the trouble. In Task 7-4-1, however, it is most efficient always to select the component requiring least work, i.e., to use the *CBA* checking sequence exclusively, since that sequence requires removal of fewer screws, on the average, than any other sequence.

For the reader who cares to verify these statements about the relative efficiency of various checking sequences we shall describe a method for calculating the mean amount of work associated with consistent use of a single checking sequence. Consider the *ABC* sequence in Task 7-4-1. With the probabilities listed in Table 1, the mean frequencies of troubles *A*, *B* and *C* in a block of six trials are 3, 2, and 1, respectively. Since we are considering a se-

TABLE 1

RELATIVE FREQUENCIES OF TROUBLE, AND AMOUNT OF WORK (NUMBER OF SCREWS) ASSOCIATED WITH THE THREE EQUIPMENT COMPONENTS

Component	Task 3-2-1		Task 7-4-1	
	Probability of Trouble	Number of Screws	Probability of Trouble	Number of Screws
A	.500	3	.500	7
B	.333	2	.333	4
C	.167	1	.167	1

quence in which *A* is always checked first, exactly seven screws would be removed on each of the three occasions when *A* is the trouble source, or a total of $7 \times 3 = 21$ screws. On the two occasions when *A* is not faulty but *B* is, each check would require removal of $7 + 4 = 11$ screws, or a total of $11 \times 2 = 22$ screws. On the one occasion when *C* is faulty, $7 + 4 + 1 = 12$ screws would be removed. The total number of screws removed in six trouble occurrences would be $21 + 22 + 12 = 55$, an average of 9.17 screws per identification of the trouble.

It will be noted that the *same* probabilities were used for both tasks, but that the amounts of work *differed* in the two tasks and that the task designations "3-2-1" and "7-4-1" refer to these differences in amount of work. It was possible to use the same trouble sequences for both tasks, resulting in treatments that differed only with respect to the work variable. Each *S* worked at only one of these tasks.

It would have been possible to randomize the troubles in blocks of six trials, with section *A* containing the trouble on three of the six trials, *B* on two other trials, and *C* on the remaining trial. However, in preliminary experimentation, such a procedure proved to be an inadequate simulation of independent probabilities, since one *S* kept an exact count of the frequency of trouble in each section, noted that after every sixth trial the ratios were 3-2-1, and concluded that six-trial blocks were being used. Therefore he was able to be certain which section contained the trouble on at least the sixth trial of each block, and, occasionally, on the last two or even three trials of the block. For this reason a more complex design was devised which maintained the desired relative frequencies over the 48-trial blocks, while eliminating the more frequent cyclic regularities.

The sequences in which the three troubles occurred were determined with the aid of a table of random numbers. Each of 12 sequences of 192 trials was used for one *S* in Group 3-2-1 and one *S* in Group 7-4-1.

For any *S*, alternatives *A*, *B* and *C*—the alternatives with trouble probabilities of .50, .33, and .17, respectively—were kept in the same positions in the equipment throughout the 192 trials. However, all six possible arrangements of the three alternatives were used in the experiment, with each arrangement being used for two *S*s in each group.

Results

Analyses of Preferences

First choice preferences.—Table 2 shows, as a function of blocks of 12 trials, the relative frequency with which each alternative was the first to be checked. In both groups during the early trials the preferences were ordered in agreement with the number of screws that had to be removed. Alternative C (low work—low probability) was the favorite and A (high work—high probability) was least favored. This preference was more marked in Group 7-4-1, where the differences in the amounts of work were greater among alternatives. Preference for the low work—low probability alternative at the beginning is accounted for by the fact that differences in amount of work could be immediately observed, while information about trouble probabilities could only be accumulated over a series of trials.

TABLE 2

Relative Frequency with Which Each Alternative Was First to be Checked

12 Trial Blocks	Task 3-2-1 Alternatives			Task 7-4-1 Alternatives		
	A (3)	B (2)	C (1)	A (7)	B (4)	C (1)
1	.264	.319	.417	.181	.257	.563
2	.403	.306	.292	.285	.306	.410
3	.500	.271	.229	.313	.306	.382
4	.556	.215	.229	.313	.292	.396
5	.569	.313	.118	.389	.243	.368
6	.556	.201	.243	.361	.271	.368
7	.597	.236	.167	.306	.257	.438
8	.660	.194	.146	.347	.215	.438
9	.618	.271	.111	.375	.264	.361
10	.576	.319	.104	.389	.250	.361
11	.653	.250	.097	.396	.257	.347
12	.632	.257	.111	.396	.181	.424
13	.618	.271	.111	.347	.194	.458
14	.618	.250	.132	.292	.187	.521
15	.667	.229	.104	.354	.174	.472
16	.632	.292	.076	.375	.208	.417

In Group 3-2-1 these preferences quickly reversed, and on the final block of 48 trials Alternative A was *S*s' first choice 63.4% of the time, B 26.0% and C 10.6%. This is in agreement with the conclusion of the earlier experiment that when all checking sequences are equally effi-

cient, *Ss* tend increasingly to select that one of the previously unchecked alternatives which has the greatest probability of uncovering the trouble, even though that alternative also requires the most work.

In Group 7-4-1 the first choice preference for Alternative C decreased and that for A increased, with A and C becoming co-favorites over the middle blocks of trials, but with Alternative C possibly regaining some of its advantage near the end of the experiment. Over the final block of 48 trials Alternative C was the first choice 46.9% of the time, A 33.7%, and B 19.4%.

It is clear that changing the work loads so as to increase the efficiency of low-probability choices and decrease the efficiency of high-probability choices resulted in parallel changes in the frequency of occurrence of these choices. However, it is equally clear that *Ss* did not adopt the most efficient strategy of *always* selecting the low probability—low work alternative, for in the final 48 trials this low probability—low work alternative was *Ss'* first choice only 46.9% of the time.

Second choice preferences.—On many trials the first choice proved incorrect, making it necessary for *S* to make a second selection. Of the two alternatives between which the second choice was made, one had a higher probability of containing the trouble but also required more work than the other. Table 3 shows the frequency with which *Ss'* second choice was (*a*) the high-probability alternative, and (*b*) the low-probability alternative. Also shown is the *relative* frequency of high-probability choices.

In the beginning, as with the first choice, both groups preferred the low probability—low work alternative to the high probability—high work alternative. The preference of Group 3-2-1 quickly reversed; after 24 trials these *Ss* consistently preferred the high-probability alternative, and, in fact, on the final 48 trials they chose

TABLE 3

FREQUENCY WITH WHICH HIGH-PROBABILITY AND LOW-PROBABILITY ALTERNATIVES WERE SELECTED AS THE SECOND CHOICE

12-Trial Blocks	Task 3-2-1			Task 4-7-1		
	Frequency of Choice		Proportion of High p Choices	Frequency of Choice		Proportion of High p Choices
	High p	Low p		High p	Low p	
1	36	62	.367	26	81	.243
2	42	57	.424	31	73	.298
3	60	41	.594	41	51	.446
4	55	35	.611	52	49	.515
5	55	33	.625	45	51	.469
6	51	34	.600	41	52	.441
7	49	35	.583	52	51	.505
8	47	37	.560	50	45	.526
9	59	34	.634	50	60	.455
10	65	33	.660	46	55	.455
11	54	37	.593	48	53	.475
12	55	25	.688	48	48	.500
13	61	20	.753	48	48	.500
14	56	30	.651	51	48	.515
15	52	39	.571	51	49	.510
16	50	25	.667	40	48	.455

TABLE 4

PROPORTION OF OCCASIONS, FOLLOWING INCORRECT SELECTION OF A, B, AND C, ON WHICH *Ss* CHOSE THE HIGH-PROBABILITY ALTERNATIVE

24-Trial Block	Task 3-2-1			Task 7-4-1		
	Incorrect First Choice			Incorrect First Choice		
	A (3)	B (2)	C (1)	A (7)	B (4)	C (1)
1	.489	.463	.289	.485	.207	.242
2	.711	.615	.429	.733	.434	.389
3	.780	.522	.400	.588	.417	.400
4	.719	.513	.300	.620	.362	.535
5	.708	.61.	.517	.569	.375	.422
6	.742	.547	.440	.581	.415	.457
7	.722	.651	.562	.636	.436	.482
8	.698	.510	.476	.560	.464	.455

it 65.8% of the time. The preference of Group 7-4-1 for the low-probability alternative also disappeared by the end of the first 48-trial block, but only to the point of equal proportions of high- and low-probability selections throughout the 16 blocks of trials. This result is especially interesting in Group 7-4-1 since it indicates that increasing experience led to decreased preference for the more efficient choice.

Each value in Table 3 is a composite of three separate values representing *Ss*' second choices following incorrect first choices of A, B, and C, respectively. In Table 4 this breakdown is presented for both groups. It shows the proportion of occasions, following incorrect selection of A, B, and C, respectively, on which *Ss* chose that one of the two remaining alternatives which was more likely to contain the trouble. It is apparent from both sets of data that the relative frequency of the high-probability alternative on the second choice is not independent of *Ss*' first choice. In Task 3-2-1, the higher the probability of success associated with *Ss*' first choice on a trial, the more likely he was to select the high-probability alternative if a second choice was required. Similarly, in Task 7-4-1, when *Ss*' first choice was A, he also preferred the high-probability alternative when a second choice was needed. However, when *Ss*' first choice was either B or C, he preferred the low-probability alternative as a second choice, though this preference seems to have been dissipating toward the end of the experiment.

Again the data for Group 3-2-1 in Tables 3 and 4 are consistent with the conclusion of the earlier experiment; namely, that given equally efficient sequences for checking the several alternatives, *Ss* tend to select that one of the remaining alternatives which has the highest probability of containing the trouble and which requires the most work. Also, the data for Group 7-4-1 in Tables 3 and 4 indicate that increasing the work associated with the high-probability alternatives decreases the popularity of the high-probability choice. However, despite the fact that the high-probability choice resulted in the greatest average amount of work, it continued to be made about as frequently as the choice with the low probability of trouble. This again shows that the probability feature of the task is weighted more heavily as a behavior determiner than is the work factor.

Verbalized preferences after 192 trials.—In addition to noting the choice sequences that occurred during the experiment, *E* asked each *S* at the conclusion of the experiment which checking sequence he would use in order to do a minimum amount of work if he were given another set of these devices to repair and if he were

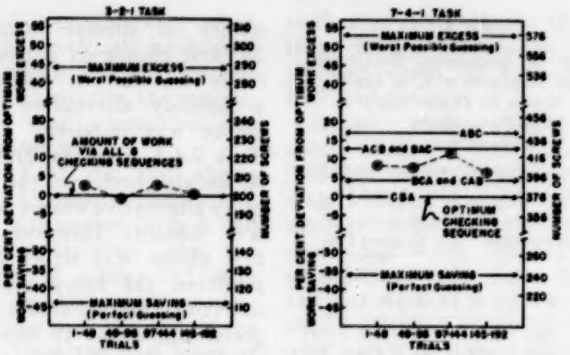


FIG. 1. Mean amounts of work that (*a*) actually were performed by *S*s per block of 48 trials (circles joined by broken lines), and (*b*) would have been performed had *S*s used the checking sequences indicated (solid lines with arrow heads).

required always to use the same sequence. The number of *S*s choosing each of the six possible sequences is shown in Table 5. For Task 3-2-1, in which all sequences required the same amount of work, 9 of 12 *S*s said they would use the ABC sequence. Seven of 12 *S*s in Group 7-4-1 also said they would use the ABC sequence, the sequence which in this task would result in an average of 17% more work than the optimum CBA sequence. This optimum sequence, on the other hand, was preferred by only one *S*, an *S* who during the entire 192 trials never checked in any sequence except C first, then B, then A.

TABLE 5

NUMBER OF *S*s WHO CHOSE EACH SEQUENCE WHEN ASKED WHICH ONE THEY WOULD USE TO MINIMIZE WORK

Task	Sequence					
	ABC	ACB	BAC	BCA	CAB	CBA
3-2-1	9	1	1	0	0	1
7-4-1	7	1	0	0	3	1

Analyses of Work

The remaining data are concerned with the amount of work *S*s did. The measure of work was the number of screws removed while searching for the trouble. The minimum possible is the number that would be removed if *S*'s first choice were correct on every trial. Similarly, the maximum is the number that would be removed if *S*'s first and second choices were incorrect on every trial, so that he removed all of the screws on every trial. Figure 1 shows these limits for both groups.

Mean amounts of work.—Figure 1 also shows the amount of work per *S* that occurred during each block of 48 trials and the amount of work associated with consistent use of any of the 6 checking sequences. Two ordinates are provided, one for the number of screws removed and the other for the percentage by which this number deviates from the number that would have been removed through consistent use of the optimum checking sequence.

In Task 3-2-1 all six checking sequences are equally efficient, each requiring removal of 200 screws in a

48-trial block. Therefore, in this task the deviation from optimum is the number of screws actually removed minus 200, and the percentage deviation from optimum is this difference divided by 200.

Since in Task 3-2-1 the amount of work is statistically independent of the checking sequence used, the expected deviation from optimum is zero, and as the four plotted points indicate, the obtained work deviation was quite close to this expected value. Over all trials the group did only 1.1% excess work.

In Task 7-4-1, the CBA checking sequence is optimum since its use results in removal of fewer screws, on the average, than use of any other checking sequence. Because all other sequences require more work, the heavier weighting of the probability variable, which resulted in occurrence of other than CBA sequences, should have produced excess work. That it did is evident in Fig. 1. The group did 6.5% excess work on the final block of 48 trials, and 8.5% excess work over all trials.

Distributions of amounts of work.— In the two groups only one *S* consistently used a single checking sequence. All others used mixed strategies, producing distributions of amounts of work about the group means. The black dots in Fig. 2 are the distribution of amounts of work in Group 3-2-1 for the final 48 trials. These points were obtained by computing each *S*'s deviation from optimum and then ordering these values from the most negative to the most positive. Thus each plotted point represents both the amount of work performed by one *S* (expressed in terms of the percentage by which it deviates from optimum) and the rank order of that amount. The projection of any of these points on both the ordinate and abscissa indicates the proportion of *S*s who did the indicated amount of work or less.

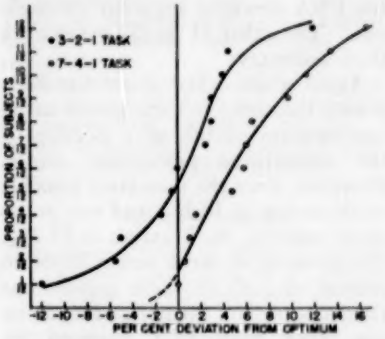


FIG. 2. Proportion of *S*s whose amount of work on the final 48 trials deviated from "optimum" by no more than the indicated amount.

A freehand curve was drawn through these points, and if we could appropriately assign values along the ordinate, this curve could be regarded as an estimate of the proportion of *S*s in the population represented who would do no more than the indicated amount of work. However, this has not been done because assigning such values on the basis of so small a sample is subject to considerable error.

In turn, the derivative of such a cumulative probability curve would be an estimate of the probability distribution of work deviations from optimum associated with Task 3-2-1. While such a curve is not presented, the data points suggest that it would be an approximately bell-shaped curve with mean near zero. The tails should extend to 44%, the maximum deviations possible under this Task.

Similar computations were made for *S*s in Group 7-4-1, and after rank ordering, the data points were plotted as circles in Fig. 2. This plot shows that only one man did *no* excess work. He, of course, was the man who used

the CBA checking sequence throughout. The other 11 *S*s did more work than necessary.

Again a smoothed curve has been drawn through the data points as an approximate sketch of a portion of the cumulative probability curve. However, since the maximum possible work saving is 36.2% and the maximum possible work excess is 53.2%, the theoretical curve would have to extend to both of these points, but the proportion of the population doing less work than that required by consistent use of the optimum checking sequence would undoubtedly be quite small.

Discussion

This experiment has confirmed a conclusion drawn in a previous study, namely, that *S*'s behavior in some multichoice situations is influenced more by the probabilities of success associated with the alternatives among which he must choose than by the amounts of work associated with these alternatives (1). Among the factors that appear to be determiners of the class of situations for which this conclusion holds are (*a*) the absolute amount of work required and (*b*) the correlation between energy demands and work times. In both the present study and the earlier one the absolute amounts of work were quite small. In addition, the energy demands were highly and positively correlated since both were largely the result of performing a single repetitive activity. Assuming that two screws require, on the average, twice the number of repetitive movements needed to remove one screw, it is apparent that (*a*) both the time and energy requirements increased as the number of screws increased, and (*b*) the energy requirement per unit of time remained approximately constant. Many other time-energy relationships are possible including high negative correlation; however, until information is available concerning the interaction of probabilities, work times, and energy requirements as reflected in behavior it would seem to be appropriate to discuss the present results in terms of the "amount of work" required, while keeping in mind that the situation had co-varying components.

One of the results of the way in which probability and amount of work jointly influence behavior is that in a large number of situations people will do more work than is necessary. In any situation for which there is an optimum checking sequence, use of any other sequence will, on the average, result in excess work. Thus either consistent use of a single nonoptimum checking sequence or mixed strategies such as were employed by all but one of the *S*s will result in excess work.

The results of these experiments have potential usefulness to both the men who work in similar work-probability situations and to their employers. In Task 7-4-1, in which the poorest checking sequence required 17% more work than the best, the *S*s did 8.5% excess work. Had each *S* used the single checking sequence he reported at the end of the experiment to be best, the group would have done 12.5% excess work. Since men apparently do not discover the efficient procedure as a result of performing such tasks, it would seem worth while in certain instances to take steps to assure more nearly optimum performance. One promising possibility would require the keeping of records on (*a*) symptoms that occur, (*b*) troubles that occur in conjunction with each symptom, and (*c*) amounts of work involved in locating and repairing the various troubles (see 2). Given sufficient information of this sort, it would be possible to estimate mathematically, for each symptom, the relative efficiency of all work sequences, and then to designate the most efficient sequence as standard procedure.

Summary

This study is concerned with the effects which certain features of problem-solving tasks have upon behavior. The type of problem-solving situation studied is one in which an indication

of malfunction has a probabilistic relationship to several possible defective components, each of which differs from the other in the probability with which it is defective and in the amount of work that is required to check it.

The results indicate that (*a*) the probability variable weighed more heavily than the work variable in determining *S*s' choices, (*b*) the increase in the efficiency of low-work choices and the decrease in the efficiency of high-probability choices resulted in parallel changes in the frequency of occurrence of these choices, and (*c*) while the low-work choice became more popular when it was made the most efficient choice, 11 of 12 *S*s did not adopt the simple, most efficient strategy of *always* selecting the low-work alternative.

The *S*s' use of mixed strategies in the task where it was most efficient always to select the low-work alternative resulted in that group doing 8.5% excess work. Only one *S* in that group correctly identified the optimum checking sequence, while 7 of the 12 *S*s actually named the *least* effective sequence as best. Had each *S* used only the checking sequence he reported to be best, the group would have done 12.5% excess work.

The results suggest that in many tasks men do not discover efficient behavior sequences as a consequence of performing the task. A method was outlined for achieving more nearly optimum performance in such situations.

References

1. Detambel, M. H. Probabilities of success and amounts of work in a multichoice situation. *J. exp. Psychol.*, 1956, **51**, 41–44.
2. Stolurow, L. M., Bergum, B., Hodgson, T., & Silva, J. The efficient course of action in "trouble shooting" as a joint function of probability and cost. *Educ. Psychol. Measmt.*, 1955, **15**, 462–477.

(Received January 9, 1956)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

RELATIONS OF CREATIVE RESPONSES TO WORKING TIME AND INSTRUCTIONS[1]

PAUL R. CHRISTENSEN, J. P. GUILFORD, AND R. C. WILSON[2]

University of Southern California

Most of the tests used in a recent factor analysis of creative-thinking abilities (**7**) are open-end tests that permit the examinee to write as many ideas or associations as he can in the allotted time. Among the scores obtained from these tests are those for fluency and originality. The fluency score in a particular test is simply the number of responses produced. Originality scores are derived from three different indicators: cleverness, remoteness of association, and uncommonness of association. In factor analyses, all three kinds of originality scores have appeared together projected on a factor vector that can be interpreted as *originality* (**6, 7**).

In an attempt to achieve further insight as to what goes into these scores of fluency and originality, an experimental analysis was made of certain attributes of the responses in relation to time during the production period. It was expected that the results would lead to ways in which measures of both fluency and originality could be improved.

The number of recalled responses produced as a function of time has been investigated by Bousfield and Sedgewick (**2**) and by Johnson, Johnson, and Mark (**5**). Using tests that emphasize simple recall (such as naming "all the U. S. cities you can think of"), they found that the rate of producing responses was relatively high during the first part of the work period but became progressively lower as time went by.

Hypotheses.—In the present study it was hypothesized that for tests that emphasize inventiveness rather than simple recall the rate of production would be relatively constant. Further, it was hypothesized that the more original responses would come later in the production period. In the case of one aspect of originality, that of cleverness, it was of additional interest to examine relative effects of instructions to write clever responses vs. instructions to write simply appropriate responses. We expected that instructions to be clever would reduce the total quantity of output of responses and raise the proportion of clever responses, but we had no hypothesis regarding the effect upon quantity of clever responses.

General Procedure

Samples tested were Air Force aviation cadets, student officers, and college students. For some of the tests used in this study, measures of time were obtained by asking *Ss* to draw lines at 2-min. intervals under the last written responses. These tests are Plot Titles (clever instructions), Plot Titles (free instructions), Impossibilities, and Brick Uses. The first three were timed over a 12-min. period and the last over a 16-min. period. Other tests used in the study were Unusual Uses, Figure Concepts, and Number Associations. Further details of procedure will be described in connection with the report of each part of the study.

[1] Based on a paper presented at the annual convention of the American Psychological Association in Cleveland, September 8, 1953. The work was done under Contract N6onr-23810, Project 150-044, with the Personnel and Training Branch of the Office of Naval Research. The ideas expressed in this paper are our own and do not necessarily represent the opinions of the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at Reed College.

Results

Relation of Number of Responses to Time

The first results to be mentioned concern the number of responses written in the successive time increments. The four creative-thinking tests involved are labeled 1, 2, 4, and 5 in Fig. 1. In the Brick Uses test (Test 5) *S* writes all the uses he can think of for a brick in 16 min. Test 4 (Impossibilities) asks for all the impossible things that can be thought of in 12 min. Tests 1 and 2 present two brief stories. For each story *S* is allowed 12 min. in which to write as many appropriate titles as he can. In Test 2 the instructions are simply to write relevant titles. In Test 1 the instructions are to write titles that are both relevant and clever.

Results

In Fig. 1 are given seven of the curves found by Bousfield *et al.* (**1, 2**) as well as the curves for the creative-thinking tests just described. A point on any curve shows the average number of responses produced up to that time. The list of tests at the top of the figure is in the rank order of the relative degree of curvature in the curves, those at the beginning of the list having the least curvature. This ranking is in the order of Bousfield's "m" constant, which can be interpreted as the rate at which a curve approaches its asymptote.

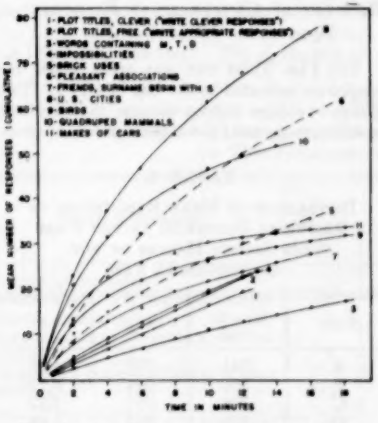


FIG. 1. Production curves of cumulative number of responses as a function of time.

With the verbatim-recall tasks (such as listing makes of cars, birds, or cities) the rate of production is rapid at first but progressively slower with elapsed time. Production curves of this kind are described mathematically by Bousfield, based upon the principle that the rate of production of words at any moment is proportional to the number remaining available.

For the more inventive or creative tasks there is decidedly less decrement in the number of responses produced toward the end of the time period than for verbatim recall tasks. It appears that in tests where (in contrast to simple recall) the task involves more imaginative work such as producing plot titles or naming impossibilities, the rate does not tend to vary with time, within the limits of the experimental working times.

Relation of Uncommonness of Responses to Time

Bousfield and Barkley (**1**) found that in recall tests the most commonly occurring words are written in the earlier stages of the production period. Their results were obtained from three subordinate-naming tasks, which called for the naming of birds, tools, and celestial bodies.

In Fig. 2 are curves from three of the creative-thinking tests. The *order* in which the responses were written is plotted on the *X* axis and the degree of uncommonness of occurrence in the group on the *Y* axis. In this study the sample was composed of approximately 400 aviation cadets and student officers.

The uncommonness values were derived empirically by determining the frequency of occurrence of every response within the group (**6**). Each response was rated on a scale from 1 to 5. Responses rated 1 were among the most common in the group, and those rated 5 were among the least common responses.

Results

Each curve in Fig. 2 represents averages for 50 *S*s, from among those

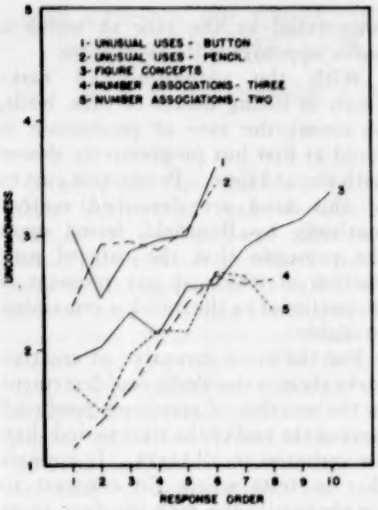


FIG. 2. Mean ratings of uncommonness vs. response order.

Ss who wrote at least 6 responses (in the case of Curves 1 and 2), at least 10 responses (in the case of Curve 3), and at least 8 responses for the remaining two curves.

Curves 1 and 2 represent two items from the Unusual Uses test, which ask for six uses for which the object or parts of the object could serve.

Curve 3 represents the data from the Figure Concepts test, which consists of 20 simple line drawings of objects and individuals. The task is to find common features that are suggested by two or more drawings and to list these features.

Curves 4 and 5 are from the Number Associations test, in which *S* is asked to write synonyms, common uses, or things commonly associated with the given numbers.

In spite of the irregularities in the curves the relationship between time and uncommonness is quite clearly a positive one.

Relation of Remoteness of Responses to Temporal Order

A second approach to originality studied in its relation to time depends upon the degree of remoteness or indirectness of the response. For example, one of the items in the Consequences test asks "What would be the consequences if everyone in the world suddenly went blind?" Direct associations are responses such as "There would be much stumbling" or "It would be dark." Examples of more remote or indirect associations are "Those who had been blind before would become the new leaders" or "New modes of transportation would be developed."

In scoring the test, responses were assigned weights of 1, 2, or 3, depending upon degree of remoteness. One hundred of the test papers from the cadet population were used to determine whether or not remoteness of connection increases with time. A mean weight was found for the first half of the written responses and a second mean weight for the second half of the responses for each *S*. These means were averaged to provide a first-half mean for the 100 *Ss* and a second-half mean.

Four items of the Consequences test were treated in this way with a t ratio computed between the two halves in each case.

Results

Three of the mean differences were found to be significant at the .01 level and the fourth at the .05 level (see Table 1).

Relation of Cleverness of Responses to Time

The Plot Titles test was given under the conditions indicated in the box in Fig. 3. Two groups of college undergraduates, 26 individuals in each group, were given both stories or plots.

TABLE 1

DIFFERENCE IN MEAN REMOTENESS OF RESPONSES OCCURRING IN THE FIRST AND SECOND HALVES OF THE CONSEQUENCES TEST

Item	$M_2 - M_1$ (d_M)	σ_{d_M}	t
A	.281	.057	4.93
B	.366	.054	6.77
C	.208	.059	3.52
D	.129	.052	2.48

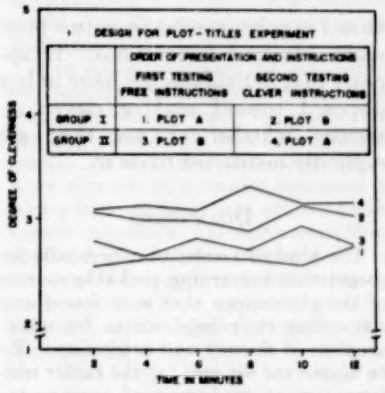


FIG. 3. Mean cleverness ratings of responses as a function of time.

First everyone was asked to write as many appropriate titles as he could for the plot. For the succeeding plot everyone was asked to write clever titles, i.e., titles that were "interest-arousing, catchy, inviting, or novel," in addition to being appropriate. The order of presentation of the two story plots was reversed for the two groups as indicated in Fig. 3. The order for the two instructions, naturally, could not be effectively reversed also.

More than 1,000 different titles were obtained for each story (under both instructions), each of which was typed on a separate slip of paper. Duplicate responses were not included. Two judges (not the authors) independently sorted the responses into five ranked categories of cleverness, achieving a normal distribution. A value of 5 indicated highly clever responses while 1 indicated nonclever. The coefficients of correlation between the ratings of these two judges were .50 for both plots.

Correlations with three other judges who rated 100 selected responses indicated that the two judges who rated all the responses were sufficiently representative for the purposes of this study so that their mean rating could be used. Since the time interval in which each response occurred was known to the authors, it was possible to obtain a mean cleverness rating for the responses of each examinee by time period.

Results

It would appear from the graph in Fig. 3 that the predicted increase of cleverness with time did not occur. In the analysis of variance applied to the data, the variation with time was found to be insignificant. Variation between plots was also insignificant.

Relation of Cleverness of Responses to Type of Instructions

As explained in the previous section, responses under the two conditions of the Plot Titles test were rated on a five-point scale of cleverness.

Results

Variation in average degree of cleverness per response due to the two different sets of instructions gives an F significant beyond the .01 point in the case of curves for Plot A, and an F significant beyond the .05 point for

TABLE 2

ANALYSES OF VARIANCE FOR DEGREE OF CLEVERNESS PER RESPONSE ON TWO FORMS OF THE PLOT TITLES TEST

Source of Variation	Plot A			Plot B		
	df	MS	*F*	*df*	MS	*F*
Between instructions	1	22.56	24.86**	1	5.93	6.48*
Between subjects in the same group	50	.91	—	50	.91	—
Between time intervals	5	.06	.38	5	.16	.91
Interaction: time intervals × instructions	5	.25	1.60	5	.07	.38
Interaction: pooled subjects × time intervals	250	.16	—	250	.18	—

*P = .05.
**P = .01.

Plot B (see Table 2). Instructions appear to make a more pronounced difference in the cleverness for Plot A than for Plot B. Plot A presents a more concrete situation and contains within it more elements that can be twisted or combined in novel ways, and it may thus provide a greater opportunity for making deliberately clever responses.

As expected, fewer total responses are produced when respondents are instructed to be clever, as shown in Fig. 1 for the two Plot Titles tests. The difference between the totals after 12 min. is significant beyond the .01 level.

Since more responses were produced under "free" instructions, there was the possibility that although the average *degree* of cleverness per response was higher under clever instructions, the total *number* of clever responses might be about the same under the two sets of instructions.

Means and *SD*s for the scores from the two plots were equated by linear transformation and a *t* test was applied, taking into account the correlation between the number of clever responses produced under the two sets of instructions. Responses were considered clever if they received an average rating of 2.5 or higher on the 5-point scale of cleverness. The resulting *t* was 3.98, which is significant well beyond the .01 level. A significantly greater number of clever responses was produced under the "clever" instructions even though there was a smaller average total number of responses produced under this condition.

The correlation between the *number* of clever responses under "free" instructions and under "clever" instructions was .62. This value corresponds favorably with reliability coefficients of cleverness scores obtained in other studies from two plots under identical instructions. It appears that individuals are more or less disposed toward making clever responses whether or not they are explicitly instructed to do so.

Discussion

The kind of results obtained calls for speculation concerning probable sources of the phenomena that were found and concerning their implications for measurement of fluency and originality. To be accounted for are: (*a*) the rather uniform rate of production of creative responses in contrast to the negatively accelerated rate of production of simple recall responses; (*b*) the increasing uncommonness of responses and remoteness of associations as time passes in the working period; and (*c*) the uniform level of cleverness of responses over the working time. There are also the findings that under instructions to give clever responses the effect is to decrease total quantity but to increase the number of clever responses and the average degree of cleverness. The third effect is, of course, related to the first two.

If we compare the task of simple recall of objects in a specified category with that of producing new ideas, we may describe the crucial difference in terms of the directness of the learned connections. That is, in the simple recall task all of the responses have been learned in direct connection with the stimulus, but in the production task the connections operate only by way of transfer of learning. Thus, in the latter case, the associations are indirect.

It can be hypothesized that the more direct the associations the more rapidly will they function in producing responses to stimuli. It is very likely that even in the simple recall situation, associative connections differ from one another in strength because of the ordinary conditions of frequency, recency, and so on. It could be stated as one hypothesis that a skewed distribution of strengths is responsible for the negative acceleration

in production, in place of or in addition to the fact that an upper limit is approached.

If an upper limit were the only reason for negative acceleration, it would seem that the curves for the productive-thinking tasks would also show the same feature. There may be, after all, a finite number of potential responses to such stimuli as have been utilized in the tests of this study. It seems more likely that the uniform rate of production of responses is due to the fact that the remote associations are, for practical purposes, of about equal strength. The curvature in the production function for the Brick Uses test is an exception. Here there are some rather familiar uses that can be produced early by simple recall. The fact of generally slower rates of production for the other four creative-thinking tests supports the hypothesis of the operation of only tenuously learned connection in those tasks.

The linear functions might be due in part to extremely large supplies of potential responses, of course. The approach to such a limit at some remote point in time might not give rise to a noticeable departure from linearity within the first few minutes of working time. It would be of interest to determine how long the same rates of production would continue, if fatigue effects could be kept out of the picture.

It was expected that, in general, later responses would be more original than earlier ones. This proved to be the case when uncommonness and remoteness of association are the criteria of originality but not when degree of cleverness is the criterion. Both uncommon and remote responses should have relatively low associative strengths and their lateness in a series of responses is therefore natural. In spite of the fact that we have demonstrated that the kind of person who gives a larger number of uncommon or remotely associated responses is also likely to give more clever responses, in the single individual the relationship no longer seems to hold in terms of their relations to response time. This might suggest that clever responses are on the average just as closely associated with a stimulus as are non-clever responses. Whether they are just as common as nonclever responses is something that could possibly be determined by actual counts. The results indicate indirectly that they may be just as common, but a study on this point still remains to be done.

The decrease in quantity of total output under instructions to be clever is probably due to the self-imposed censorship of *S* under those instructions. Either the instruction inhibits the calling up of some of *S*'s responses entirely or it prevents his writing responses that he rejects as nonclever. It would be a good hypothesis that both of these things occur. That the average level of cleverness is increased under instructions to be clever is an indication that this is something that can be achieved voluntarily, to some extent, at least under the circumstances of the Plot Titles test. Whether instruction to be clever induces individuals to exceed their upper limits of cleverness under the "free" instructions, is not known.

It will be noted that the temporal relationships examined were studied by using different tests for the three different types of scores—uncommonness, remoteness, and cleverness—rather than by using the same test for the different scores. It is believed, however, that since all three types of scores have been shown to be related to the factor of *originality*, the findings nevertheless have generality. It happens that each type of score is more naturally and conveniently derivable from certain kinds of tests.

One or two implications for testing practices may be seen in the results. In the measurement of fluency, it seems that testing efficiency continues at about the same level in most tests over a period of 12 min. per stimulus. This gives considerable freedom of choice between using the total time on a few items with a longer time per item vs. a larger number of items with a shorter time per item. This conclusion is based upon informa-

tion regarding means, assuming that variabilities are proportional to means.

The implications for the measurement of originality are not so clear, although the indication is that longer times per item are more favorable to bringing out variance in remotely associated and uncommon responses.

The finding that fewer but relatively more clever responses are produced under instructions to be clever indicates some advantage in establishing a particular set for a group. Under the special instructions to be clever, the associative process would seem to be restricted and redirected through self-criticism. Where a score for fluency is derived from the number of nonclever responses, however, these more restrictive instructions may be less desirable.

That differential, self-imposed sets are also operating is suggested by the finding that the same individuals tend to produce the more clever responses regardless of instructions. The substantial correlation between scores under the two conditions would indicate that for selecting the more clever individuals one set of instructions would do about as well as the other. The results of the present study tell us nothing about the effects upon factor loadings. We have results from two previous studies [where the instructions were different in the two studies, (3, 4)], however, indicating that there is no appreciable difference in the *originality* loading of the tests under the two sets of instructions.

Summary

This study explored relationships between the number of responses produced and the successive increments of time in which the responses were produced, for several creative-thinking tests. Other dimensions explored in relation to time were three aspects of originality: cleverness, uncommonness, and remoteness. In the case of cleverness, the effects of instructions to be clever were also studied.

As for the quantity of responses produced, examination of the production curves indicates a decrease in rate with time for simple recall tasks but a relatively constant rate for more inventive or creative tasks.

For the creative-thinking tests analyzed, it was found that two of the three aspects of originality increased with time. These were uncommonness and remoteness. Cleverness, as measured in this study, was found to be independent of time.

Instructions to be clever appeared to decrease the total number of responses produced but to increase the total number of clever responses and the average degree of cleverness.

Implications for the measurement of the traits of fluency and originality are indicated.

References

1. Bousfield, W. A., & Barclay, W. D. The relationship between order and frequency of occurrence of restricted associated responses. *J. exp. Psychol.*, 1950, **40**, 643–647.
2. Bousfield, W. A., & Sedgewick, C. H. W. An analysis of sequences of restricted associative responses. *J. gen. Psychol.*, 1944, **30**, 149–165.
3. Guilford, J. P., Berger, R. M., & Christensen, P. R. A factor-analytic study of planning. II. Administration of tests and analysis of results. *Rep. psychol. Lab.*, No. 12. Los Angeles: Univer. of Southern Calif., 1955.
4. Guilford, J. P., Kettner, N. W., & Christensen, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation. II. Administration of tests and analysis of results. *Rep. psychol. Lab.*, No. 16. Los Angeles: Univer. of Southern Calif., 1956.
5. Johnson, D. M., Johnson, R. C., & Mark, A. L. A mathematical analysis of verbal fluency. *J. gen. Psychol.*, 1951, **44**, 121–128.
6. Wilson, R. C., Guilford, J. P., & Christensen, P. R. The measurement of individual differences in originality. *Psychol. Bull.*, 1953, **50**, 362–370.
7. Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. A factor-analytic study of creative-thinking abilities. *Psychometrika*, 1954, **19**, 297–311.

(Received December 20, 1955)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

EFFECTS OF EXPERIENCE ON PERCEPTION OF CAUSALITY [1]

HOWARD E. GRUBER, CHARLES D. FINK, AND VERNON DAMM

University of Colorado

Recent studies of event perception (**1, 2, 3**) have had as their chief concern the specification of the stimulus variables governing such experiences as the impression of causality and the perceived continuity of movement. Michotte's studies, in particular, provide striking evidence that the impression of causality is the immediate perceptual outcome of given spatiotemporal configurations. His analysis of the variables affecting the impression of causality is largely confined to the immediately given stimulus conditions. In the present study, the effects of a prior series of judgments on a later series is investigated.

The general method was to present *S* with a stimulus event of which the first part might or might not be perceived as the cause of the second. The stimulus event was the collapse of a "bridge" following the removal of one of its vertical members. A very definite and striking impression of causality is reported if the time between the removal of this upright and the collapse of the bridge is sufficiently short. As the time interval increases, *S* reports such an impression less and less frequently. In the experiments which follow, the major dependent variable is the "temporal threshold of causality," that is, the maximum time interval at which *S* reliably reports an impression of causality. In Exp. I and II this value was determined before and after an interpolated series of judgments which, it will be seen, influenced the threshold. Experiment III deals with some cognitive activities which accompany the perception of causality.

[1] This research was supported by the Council on Creative Research of the University of Colorado.

Apparatus.—The construction of the bridge permitted the manipulation of two experimental variables: the vertical distance ("gap") between a removable post and a horizontal bar; and the time interval between the removal of the post and the falling of the bar. The bridge consisted of two vertical posts and a horizontal steel bar. At the beginning of each trial the bar was always horizontal, with one end resting on the left post. Concealed within the left post was an electromagnet holding the steel bar in a horizontal position; when its power supply was interrupted, the bar fell. The right post, mounted on rollers, was attached to a weight by means of a pulley. When a brake was released, the post moved away from the bridge. An electronic interval timer controlling the magnet was actuated at the moment the brake was released.

The height of the left post and that of the bar could be varied from 6 in. to 8.75 in. Since the height of the right post was 6 in., this permitted *E* to vary the vertical gap between the right post and the horizontal bar from zero to 2.75 in.

A screen with a rectangular window was placed in front of the apparatus so that *S* could see only the two uprights and the bar. The movable post came to rest out of *S*'s field of view.

Experiment I

The first purpose of Exp. I was to determine the temporal threshold of causality before and after an interpolated series of trials. In these interpolated trials a delay between the two parts of the stimulus event produced a definitely "not causal" impression. The second purpose was to compare judgments of causal relation in a situation where the spatial arrangement suggested a causal connection between the parts of the

stimulus event (no gap) with judgments where the spatial arrangements ran counter to such a connection (gap). There were thus four experimental groups: No Gap, Delay Trials; No Gap, Control; Gap, Delay Trials; Gap, Control.

Method

For the No Gap groups, the bar in its starting position seemed to rest on the movable post; for the Gap groups, there was a vertical gap of 2.75 in. between the post and the bar. The Delay Trials groups received 12 trials interpolated between the pretest and the posttest. In these 12 trials the delay between the removal of the post and the falling of the bar was .35, .40, or .45 sec., varied in haphazard order. In preliminary experiments these delays had far exceeded the threshold for a causal impression. The 12 interpolated trials for the Delay Trials groups took approximately 3 min. The Control groups simply rested for 3 min. between pre- and posttests.

In the pretest and posttest the temporal threshold of causality was determined from four series of judgments, using the method of limits with stimuli presented in haphazard order. In each series, the time intervals used were .00, .05, .10, .15 and .20 sec., with one of each value occurring in every group of five trials. If all "causal" responses were given in a group of five trials, *E* added trials with longer delays to that group in .05-sec. steps until one "not causal" response was given.

The *S* was seated 6 ft. from the bridge. Instructions were as follows: ". . . This bar rests on the left-hand post and is held there by a magnet. I control the magnet; or, in other words, I determine when the bar will fall. Like this (demonstrate). When I press a switch back here this post moves away. Like this (demonstrate). Now, under certain conditions, when I release the post and the bar it may look as though the removal of the post causes the bar to fall. Actually there is no causal relationship between them. . . . We merely want to determine whether or not you get the impression that the removal of the post looks as though it causes the bar to fall." Instructions were to watch the region where the end of the bar extended over the movable post, and to respond after the bar had fallen with either "causal" or "not causal." There was a ready signal before each trial. There were 10 practice trials.

The *Ss* were 42 sophomore women. There were eight *Ss* in each experimental group. In the Gap condition, however, a temporal threshold could not be determined for every *S*. This necessitated testing 10 additional *Ss* in order to get eight for each of the Gap conditions. While these 10 were not included in the experimental groups, their results will be discussed below.

Results

Each *S*'s judgments were treated as four series, each series containing one maximum delay judged "causal." The temporal threshold of causality was defined as the mean of these four maxima.

Effects of the spatial gap.—In the No Gap condition, all 16 *Ss* gave some causal and some not causal responses, permitting a threshold determination in all cases. In the Gap condition, 20 of 26 *Ss* gave some causal responses, 6 reported no causal impressions, and 1 failed to report any in the posttest. The temporal threshold was not determined for 3 *Ss* who continued to report positively with delays of over .5 sec.

The pretest results for Delay Trials and Control groups were combined, since the two groups were treated the same until after the pretest. The median temporal threshold, based on 16 *Ss* in each condition, was .095 sec. for the No Gap condition and .098 for the Gap condition, a negligible difference.

Immediately after completing his judgments, *S* was asked to rate as strong, medium, or weak the causal impressions he had reported. In the No Gap condition, 15 *Ss* rated their causal impressions as strong, and 1 as medium. In the Gap condition, 6 *Ss* rated their impressions as strong, 6 as medium, and 4 as weak. This difference between the two conditions is significant (with Yates's correction, chi square = 8.8, $df = 1$, $P < .01$).

In the Gap condition, several *Ss* reported that they had imagined the gap filled or spanned in some way.

TABLE 1

PERCENTAGE FREQUENCY OF CAUSAL RESPONSES AS A FUNCTION OF DELAY BETWEEN MOVEMENTS

Delay (Sec.)	Gap Posttest		No Gap Posttest	
	Delay Trials	Control	Delay Trials	Control
.00	68.8	68.8	100	87.5
.05	78.1	53.1	87.5	65.6
.10	37.5	31.3	81.3	68.8
.15	34.4	15.6	47.0	15.6
.20	50.0	9.4	62.5	12.5
.25	28.1	0	18.8	2.7

For example, one *S* got quite strong causal impressions, but only if she imagined a fine wire extending upwards from the movable post and supporting the bar.

Effects of delay trials.—As shown in Table 1, the effect of the delay trials was to increase the number of causal responses at all time intervals from .00 sec. to .25 sec. The median increase in temporal threshold of causality was .075 sec. for the No Gap, Delay Trials group and .020 sec. for the Gap, Delay Trials group. These differences are significant at .01 and .05 levels, respectively, as tested by Wilcoxon's method for paired replicates (**4**). In the No Gap, Control group there was no change in threshold from pretest to posttest, and in the Gap, Control group there was an insignificant median *decrease* of .010 sec.

EXPERIMENT II

The purpose of Exp. II was to compare the effects of an interpolated series of trials producing all causal responses with the effects of an interpolated series producing no causal responses. In addition, a variation in instructions was introduced to evoke more objective judgments, as against the phenomenal orientation of Exp. I.

Method

There were two experimental groups, differing only with regard to the 12 interpolated trials: the Zero Delay group and the Long Delay group, in which the time between the removal of the post and the falling of the bar was .80 sec. in the interpolated trials. In the pretest and posttest the temporal threshold was determined from one series of 12 trials, with delays from zero to .33 sec. in steps of .03 sec., and presented in haphazard order. Only the No Gap condition was employed.

The *S*s received an explanation of the apparatus, as in Exp. I, but they were required to *decide* "whether the bridge has fallen because the support has been removed or because the magnet has been released." The *S*s were 20 college sophomore women. It was possible to determine temporal thresholds for 18; although the other 2 each gave some causal responses, their judgments were too erratic to obtain thresholds with the method employed and are not included below.

Results

The threshold for each *S* was defined as the midpoint of the zone of uncertainty. The effect of the interpolated trials was to raise the threshold in the Long Delay group and to lower it in the Zero Delay group. For the Long Delay group, the mean pretest threshold was .132 sec. and the mean posttest threshold was .160 sec.; for the Zero Delay group the corresponding values were .103 sec. and .062 sec. The mean increase for the Long Delay group was not significant ($t = 1.60$), although the change was in the direction expected from Exp. I. The mean decrease in threshold for the Zero Delay group was significant ($t = 2.97$, $P < .01$); and the difference between the pretest and posttest differences of the two groups was significant ($t = 3.00$, $P < .01$).

On the whole, it made little difference if *S*s were directed to report on phenomenal appearances, as in Exp. I, or to decide on the actual physical cause of the event observed, as in Exp. II: the impression of causal

relation depended primarily on the temporal relations between the parts of the stimulus event.

Experiment III

In the previous experiments many *Ss* spontaneously reported the use of imagery. The purpose of Exp. III was to determine the frequency with which imagery was utilized.

Method

Three gap conditions were employed: Large Gap (2.75 in.), Small Gap (.25 in.), and No Gap. The instructions were the same as in Exp. I. The *Ss* were interviewed as to the use of imagery or other forms of cognitive elaboration and as to the strength of the causal impressions received. The experimental sequence for each *S* was: Large Gap, interview, Small Gap, No Gap, interview. The *Ss* were 21 college sophomores.

Results

The instructions stressed the fact that the falling of the bar was caused by *E*'s switching off the magnet, not by the removal of the post. The spatial gap further emphasized the lack of relation between the two movements. In the interview after the Large Gap condition the *Ss* were asked if they had at any time found it necessary to imagine or assume anything in order to get an impression of causality. Five *Ss* reported that they had imagined the post extended so that the bar was resting on it. Four *Ss* reported that they had simply made the assumption that the post and bar could be causally related. Two *Ss* related the post and bar by imagining the operation of some mechanical system. There were six *Ss* who reported causal impressions but reported no use of imagery or assumptions.

In the Large Gap condition the causal impressions of 9 *Ss* were rated as strong or medium and 12 as weak or none; in the No Gap condition the corresponding figures were 17 strong or medium, and 4 none. This result is similar to that obtained in Exp. I, where the comparison of Gap and No Gap conditions was based on independent groups.

Although *Ss* were not interviewed after the Small Gap condition, their comments suggest that the presence of any visible gap was the important factor for some *Ss*. For others, however, the influence of the large gap was probably greater than that of the small one.

Discussion

Temporal contiguity of the two parts of the stimulus event was sufficient to produce an impression of causal relation, even in the gap condition, which militated against a causal relation. The impression of causality was not a rational induction but an immediate perceptual experience depending primarily on temporal sequence.

The fact that the temporal threshold of causality was unaffected by the introduction of the spatial gap confirms Yela's findings in his studies of causation at a distance (**5, 6**). But it cannot be said that the gap had no effect on *Ss*' responses. The causal impressions reported were weaker in the Gap condition than in the No Gap condition. Some *Ss*, moreover, did not give any causal impressions at all in the Gap condition.

Under the most favorable stimulus conditions the impression of causality was a direct perceptual experience, as described by Michotte. Under less favorable stimulus conditions the impression of causality did not necessarily disappear. When a spatial or temporal gap was introduced in the stimulus event, supplementary cognitive activity was spontaneously reported by many *Ss*. Spatial gaps were filled by imagining that the stimulus objects were extended in various ways. Temporal gaps were filled by imagining a new event intervening between the parts of the stimulus

event. Further research on this cognitive gap-filling activity is now in progress.

The change in temporal threshold as a result of the interpolated trials is a clear experimental demonstration of an effect of past experience on the impression of causality. The increase in threshold following the long delay trials and the decrease in threshold following the zero delay trials suggests an anchoring effect similar to that found in many perceptual experiments. Time intervals shorter than a certain critical value produce a causal impression, and longer intervals produce a not-causal impression. The threshold interval tends to shift toward the mean interval for the stimulus series. The effect of the interpolated trials was to change the mean interval.

It should be pointed out that the present study has no direct bearing on the Humean argument concerning the origin in individual learning of the category of relation we call "causality." We have demonstrated that the conditions which elicit the impression of causality can be modified by prior experience.

Summary

Three experiments on the perception of causality are reported. The stimulus event was the collapse of a bridge, with a variable time interval between the removal of an upright and the falling of a horizontal bar. The stimulus event was also varied by introducing a visible gap between the removable upright and the horizontal bar.

With no delay between the two movements, and no spatial gap, a striking impression of causality was almost invariably reported. Under less favorable conditions, an impression of causality was often accompanied by spontaneous reports of cognitive activity which served to fill the spatial gaps or the temporal delays.

The temporal threshold of causality was defined as the maximum delay at which an impression of causality would be reliably reported. This threshold was the same in the Gap and No Gap conditions. But in the Gap condition the causal impressions were weaker, and some *Ss* failed to report any causal impressions.

The temporal threshold was increased by a prior series of trials with long delays and decreased by a prior series with zero delays. This is interpreted as an anchoring effect of past trials, modifying the temporal conditions necessary to elicit an impression of causality.

References

1. Burke, L. On the tunnel effect. *Quart. J. exp. Psychol.*, 1952, **4**, 121–138.
2. Johansson, G. *Configurations in event perception.* Uppsala: Almqvist and Wiksells, 1950.
3. Michotte, A. *La perception de la causalité.* Louvain: l'Institut supérieur de Philosophie, 1946.
4. Wilcoxon, R. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1949.
5. Yela, M. An experimental study of phenomenal causation. Causation at a distance. *Quart. J. exp. Psychol.*, 1952, **4**, 139–154.
6. Yela, M. La nature du "Rayon d'action" dans l'impression de causalité mécanique. *J. Psychol. norm. path.*, 1954, **47–51**, 330–348.

(Received December 27, 1955)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

A MEASURE OF STIMULUS SIMILARITY AND ERRORS IN SOME PAIRED-ASSOCIATE LEARNING TASKS[1]

ERNST Z. ROTHKOPF[2]

Maintenance Laboratory, Air Force Personnel and Training Research Center

It is the purpose of this study to provide a measure of similarity among the signals of the International Morse Code. Such a measure is desirable for three reasons:

1. A measure of similarity[3] among Morse signals may be useful to persons concerned with the instruction of telegraphers. Several studies (**5, 9, 11**) have shown that similarity among stimulus terms determines to a large degree the difficulty of paired-associate lists. Another study (**3**) indicated that the most economical order in which a list can be presented to *S* is related to similarity among the stimuli of the list. Both learning difficulty and the most economical practice order are factors of importance in practical training situations.

2. A measure of signal similarity may increase the usefulness of aural Morse stimuli as a research tool. The task of learning to write alphabetic and numerical equivalents to Morse stimuli has much to recommend it as a vehicle for the study of theoretical problems in paired-associate learning. This task is readily fitted by the S-R paradigm. It lends itself easily to mass administration and thus makes the inclusion of a larger number of *S*s in research designs practicable. In addition, learning Morse code has probably more intrinsic motivational value for *S*s than learning the traditional paired nonsense syllables.

3. A generalizable, quantitative approach to stimulus similarity is of theoretical importance because gradients of stimulus generalization are thought to proceed along continua of stimulus similarity (**6**). It is assumed that such *hypothetical* gradients of excitation and inhibition underlie cases of empirical stimulus generalization such as substitution errors in paired-associate learning. The notion of hypothetical as well as empirical stimulus generalization has been used in theoretical accounts of a wide variety of behavioral data (e.g., **4, 6, 10**). Consequently a general measure of stimulus similarity is likely to be an important step in the construction of a theory of paired-associate learning. A similarity measure for the signals of the International Morse Code may contribute to the methodology of a more general similarity metric.

The applicability of the concept of stimulus similarity to theoretical and practical problems depends upon the method for describing similarity among stimuli. Optimally this method should be quantitative and

[1] This research was carried out as part of the United States Air Force Personnel & Training Research Program in support of Project 7714. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

[2] The author is indebted to A/1C Richard H. Perkins, A/1C Bertram B. Adams, and A/2C Arthur F. Friedman, Jr., who performed the actual experimental work and carried out considerable portions of the data analysis.

[3] Similarity is here used in the sense intended by the authors of the several reports referenced in this paragraph. Some congruence among the various definitions of similarity, including the rather narrow definition used in this paper, may be inferred because every one of these conceptualizations bears a relationship to substitution errors.

should yield a measure of stimulus similarity which bears a simple relationship to the behavioral data which it is intended to predict. An additional requirement exists if the measure is to be used in connection with paired-associate tasks. For many of these situations the method by which similarity is determined must be applicable to stimuli which vary in several physical dimensions. This is necessary because stimulus terms in practical and experimental paired-associate tasks are frequently physically complex.

Previous authors have determined "stimulus similarity" in several ways (**9,** pp. 19–21). The present paper restricts the term to measurements produced by a particular psychophysical procedure. This procedure is the method of constant stimuli. Two response categories "same" and "different" will be used. The definition of similarity proposed here is the relative frequency of "same" judgments which a pair of stimuli elicits under standard conditions described later.

The usefulness of this definition of similarity is of course dependent upon the relationship between the measure of stimulus similarity and data of interest. For practical reasons and because of the assumed relationship between similarity and empirical stimulus generalization, these data will consist for the present study of the frequency of substitution errors in a paired-associate task in which aural Morse signals serve as stimulus terms and letters or numerals serve as response terms. This report will concern itself with the measurement of similarity among the 36 common signals of the International Morse Code and with the relationship between these measures and the average number of substitution errors made to each signal during training. The observational data which were collected especially for this study include only the measures of stimulus similarity. The error data which are to be utilized were obtained from two reports in the literature of Morse code training (**8, 9**) and from two exploratory studies conducted at this laboratory in connection with other work.

Method

Procedure.—The *Ss* of this experiment were exposed to pairs of aural Morse signals sent at a high tone speed. The signals of each pair were separated by a short temporal interval. The *Ss* were asked to indicate whether they thought the two signals were the same (or different) by making the appropriate mark on an IBM True-False Answer sheet. Each *S* was asked to respond in this fashion to 351 different pairs of Morse signals.

Materials and apparatus.—Eight different lists of 351 stimulus pairs each were used in this experiment. The lists were composed in the following manner. The 36 signals of the Morse code allow 1,260 two-signal permutations to which the objectively correct response is "different" and 36 pairs of signals to which the objectively correct response is "same." The 1,260 "different" pairs were divided randomly into four groups of 315 pairs of signals each. All of the 36 "same" pairs were then added to each of the groups of 315 pairs. Each list was in this manner brought to a total of 351 pairs of signals. The entire process described above was repeated to yield four additional lists, for a total of eight lists. Pairs were then drawn at random without replacement for each list of 351 in order to determine order of presentation.

Each of the eight lists was recorded on magnetic tape. The signals were generated by a Boehme Automatic Keyer using a tone speed of 20 words per minute. This tone speed means that each signal was sent fast enough to make 20 transmissions of the word "Codex" within 1 min., or that the temporal duration of each dot was .05 sec., that each dash was .15 sec., and that a silent period between elements was .05 sec. The time interval between the signals of each pair was 1.4 sec., with 3.00 sec. between pairs. To help *S* keep place on the response sheet an interval of 4.6 sec. was interposed after every tenth pair of signals. A tape-recorded modification of the instructions used

with the Signal Corps Code Aptitude Test preceded each list.[4]

The output of a tape recorder carrying the experimental stimuli was, after further amplification, fed to loudspeakers which were mounted in the ceiling of the experimental chamber. This room was judged by *E* to have good acoustical characteristics. The signals were thought to be of comparable audibility throughout the room. Although the experiment was carried on during the month of July in southern Texas, the experimental room was not air-conditioned.

Subject and design.—The *S*s of this experiment consisted of 598 airmen awaiting basic training at Lackland AFB. The number was composed of eight separate marching units. These units ranged in size from 51 to 120 men. The *S*s who reported Morse code experience were not used. Each of the eight marching units was assigned to a different list of stimulus pairs.

Results

Stimulus similarity.—Responses to each stimulus pair were divided into three categories: (*a*) answers indicating *S* judged the two stimuli to be the same (True); (*b*) answers indicating judgment that the signals differed (False); and (*c*) answers in which neither alternative or both alternatives were marked by *S*. Only .32% of the 209,898 responses recorded during the course of this experiment belonged in the last category.

In the combined eight lists, each of the 1,260 objectively different stimulus pairs occurred twice, and each of the 36 objectively same pairs occurred 8 times. The number of "same" (True) responses for each stimulus pair were summed over all occurrences of a pair. This sum was divided by the total times that pair had been presented to *S*s. The resulting percentage of "same" judgments for each of the 1,296 stimulus pairs is available in tabular form through the American Documentation Institute.[4] This table contains a similarity measure for *each* of the *two* permutations of every pair of nonidentical stimuli. It was decided to combine these two measures because a single expression for the similarity of two stimuli simplifies subsequent computations and also avoids a difficulty of interpretation with regard to differences between judgments of a stimulus pair A:B and that of a stimulus pair B:A.[5] The measures were combined by adding the percentage of "same" responses made to each of the two permutations of a stimulus pair and dividing by two. The average percentage of "same" judgments for each stimulus pair is shown in Table 1. This is the primary measure of stimulus similarity with which this paper concerns itself.

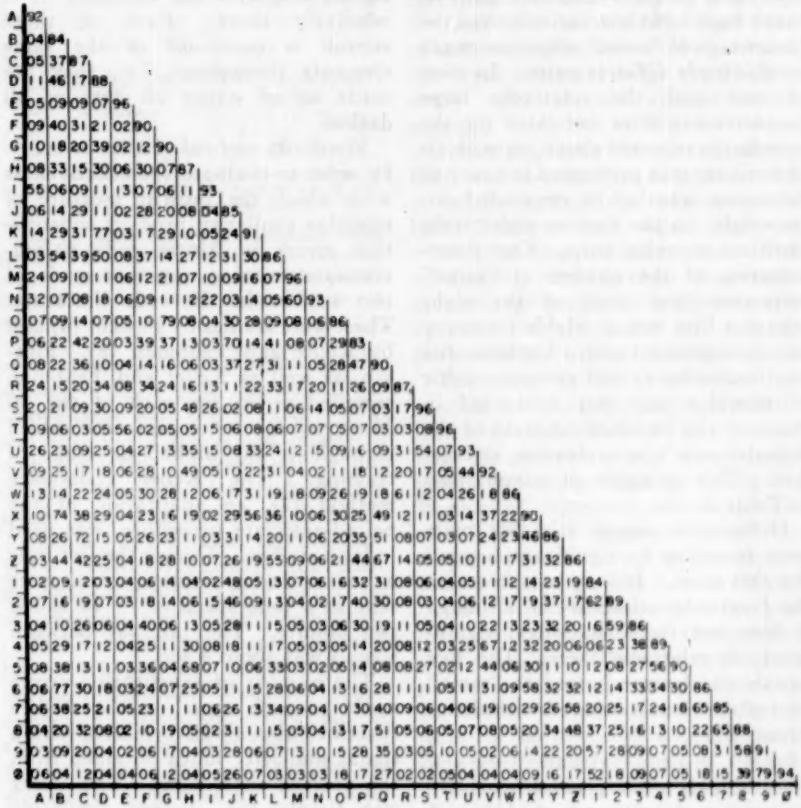
In order to get some indication of the stability of the obtained measures the percentages of "same" responses on each of two presentations of every (1,260) objectively different stimulus pair were compared. For this purpose the four lists resulting from the first drawing and the four lists resulting from the second drawing were

[4] The exact text of the experimental instructions as well as the data on which the graphs and the statistical material of this paper are based have been deposited with the American Documentation Institute. Order Document No. 5068 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] It should be pointed out that averaging the A:B and B:A cell entries is operationally equivalent to assuming that stimulus similarity is symmetric (i.e., A is as similar to B as B is to A). This assumption may limit the usefulness of the similarity measure to some extent because asymmetry with regard to substitution errors are commonly reported in the literature. For example *H* is said to be substituted for *5* more often than *5* is for *H*. Similarly *B* for *6* substitutions exceed *6* for *B* errors and so on. However the nicety of an asymmetric similarity relation between stimuli does not appear warranted in view of the large measurement error.

TABLE 1

SIMILARITY AMONG THE 630 OBJECTIVELY DIFFERENT COMBINATIONS OF COMMON MORSE SIGNALS AND PERCENTAGE OF "SAME" JUDGMENTS TO THE 36 OBJECTIVELY IDENTICAL STIMULUS PAIRS



treated as the first and second replication, respectively. In each of these replications every stimulus pair occurs once. The product-moment coefficient of correlation was then computed between the proportion of "same" judgments to corresponding stimulus pairs in the first and the second replication. The resulting r is $.784 \pm .03$. A factor which may have depressed the coefficient of correlation to some degree is a small but significant increase in the number of "different" judgments as a function of the ordinal position of the stimulus pair in the list. The method for making up the stimulus lists used in this experiment did not adequately control for this.

For computational reasons the correlation reported above was obtained only from responses to objectively different stimulus pairs. Among the population of stimuli consisting of

Morse signals sent at the tone speed used in this experiment, the percentage of "same" judgments made to *objectively identical* stimulus pairs is much higher but less variable than the percentage of "same" responses made to *objectively different* pairs. In view of this and the relatively large measurement error indicated by the correlation reported above, an analysis of variance was performed in order to determine whether *S*s responded differentially to the various objectively identical stimulus pairs. One determination of the number of "same" responses from each of the eight stimulus lists was available for every identical stimulus pair. Variance due to stimulus list as well as variance due to stimulus pair was evaluated in terms of the variance estimate of the stimulus-pair × stimulus-list interaction. This analysis is summarized in Table 2.

Differences among stimulus pairs were found to be significant beyond the .001 level. It is to be noted that the *F* ratio for stimulus list, although it does not reach statistical significance, is relatively large. The five signals which were judged as "same" *least* often when compared with themselves were: P (·––·), B (–···), 1 (·––––), J (·–––), and 7 (––···). These signals are in general relatively long and each utilizes both dots and dashes. The five objectively identical stimulus pairs which were correctly judged "same" *most* often involved the signals: E (·), T (–), M (––), S (···), and 0 (–––––). These signals are, with the exception of 0, relatively short. Each of these stimuli is composed of the same elements throughout, i.e., they are made up of either all dots or all dashes.

TABLE 2

ANALYSIS OF VARIANCE FOR JUDGMENTS OF OBJECTIVELY IDENTICAL STIMULUS PAIRS

Source	*df*	Mean Square	*F*	*P*
Stimulus pair	35	11,829.8	2.63	<.001
Stimulus list	7	8,130.3	1.80	
Interaction	245	4,505.9		
Total	287			

Similarity and substitution errors.—In order to evaluate the effectiveness with which the present measure of stimulus similarity predicts substitution errors in Morse code paired-associate training, use was made of two studies in the code literature. These were studies by Plotkin (**9**) and by Keller and Taubman (**8**). Plotkin, as well as Keller and Taubman recorded errors for each of the 36 Morse characters on test runs spaced throughout the alphabetic stage of training. The Keller code-voice method (**7**), an adaptation of the paired-associate procedure, was used in both studies. During the test runs as well as during practice, signals were sent at a tone speed of 15–18 words per minute. The interval between signals was about 4 sec.

The number of substitution errors made to each signal in the studies described above was correlated with a signal similarity index obtained from the present experiment. This index was simply the average similarity between a signal and the 35 other Morse signals. The product-moment correlation between average similarity and error measures from the Plotkin study was .74. The correlation between average signal similarity and the Keller and Taubman data was .71. The coefficient of correlation between the Plotkin and the Keller and Taubman error measures was .92. The lack of perfect relationship between the measure of similarity and the incidence of substitution errors is

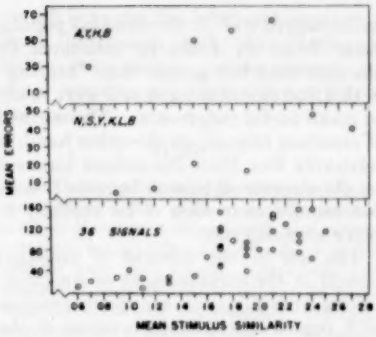


FIG. 1. Mean number of substitution errors per signal as a function of mean stimulus similarity. The 36-signal data are from Plotkin (9). The vertical scales of the three plots are not comparable.

to an appreciable degree due to the influence of a few signals. Chief among these were B (−···), C (−·−·), and 7 (−−···) which had relatively high similarity to other signals but evoked relatively few substitution errors according to the Keller and Taubman, and Plotkin reports; and W (·−−) which ranked low in average similarity but ranked very high as far as substitution errors were concerned.

Two studies conducted at this laboratory provided additional data on the relationship between the frequency of substitution errors made to signals and stimulus similarity. The first of these studies involved the signals H, A, B, and Y. Each stimulus was presented at a tone speed of 20 words per minute. The code-voice method was used. All *Ss* had 20 practice trials which brought the group average to about 50% correct responses. A 20-item test, in which each of the four stimuli occurred five times at random, was administered immediately after the last practice trial. The second study used the signals B, S, K, N, Y, and L. The training conditions were as described above, except that learning was carried to about 85% accuracy (192 trials). A 24-item test, in which each of the six stimuli occurred four times at random, was administered after the last practice trial. Error measures for each stimulus from the final tests of the two studies described above and the error measures from the Plotkin study were plotted against mean stimulus similarity in Fig. 1. Mean stimulus similarity for stimuli of the four- and the six-item lists was computed by taking the average similarity between each signal and the other (3 or 5) signals in the list. The relationship between error frequencies and average similarity is positive and appears roughly linear for a large portion of the similarity range.

The analysis of variance of responses to objectively identical stimuli which was described earlier indicated that the percentage of "same" judgments varied significantly among the 36 pairs of objectively identical stimuli. This fact suggested that the frequency with which a stimulus is judged "same" when it is presented twice in rapid succession may be a predictor of the number of errors which that stimulus is likely to evoke in a paired-associate learning situation. In order to test this notion the coefficient of correlations was computed between the percentage of "same" responses to the double presentation of each stimulus and the number of substitution error responses made to each stimulus in the Plotkin, and in the Keller and Taubman studies. These correlations were − .69 and − .70 respectively.

DISCUSSION

The data from the present study indicate that the frequency with which pairs of Morse signals were judged "same" is positively related to the number of substitution errors made be-

tween the judged stimuli in paired-associate training situations. This finding argues for the usefulness of the method of constant stimuli in the determination of stimulus similarity among signals of the International Morse Code.

The rationale for defining similarity among stimuli in terms of psychophysical judgments is based on the assumption that a psychophysical judgment is directly related to differences in the *stimulating effects* (receptor activity, afferent activity, perceptual responses, etc.) between the two stimulus energies. It is also assumed that *generalization* between two stimuli is a simple, decreasing function of the difference in their *stimulating effects*. Consequently it is to be expected that a simple relationship exists between the frequencies of certain categories of psychophysical responses to a pair of stimuli and the tendency to generalize between these stimuli. This type of behavioral law is preferable to formulations which link the physical differences between two stimulus objects and the tendency to generalize because the relation between the physical stimulus and its stimulating effect is generally complex. An often crucial, additional difficulty with the physical approach to similarity lies in the identification of physical dimensions in which complex stimuli differ from each other.

Two earlier investigators used a psychophysical approach to stimulus similarity in connection with paired-associate experiments. Both Yum (11) and Plotkin (9) employed the method of single stimuli—each investigator asked his *S*s to rate pairs of stimuli according to similarity. This technique has serious limitations when the number of stimuli among which similarity ratings are to be made is relatively large. With the 36 signals of the International Morse Code, the similarity of 1,260 pairs of stimuli must be rated (a pair of stimuli plays here the role usually held by a single stimulus). This number presents difficulties since the method of single stimuli leans heavily on *S*'s experience with the distribution and range of the to-be-judged trait in the stimulus population. Most *S*s would be exhausted by the time they had gotten their "bearing" with 1,260 stimulus pairs and were ready to make useful judgments. The method of constant stimuli, on the other hand, is relatively free from limitations imposed by the number of stimuli because it does not require that each *S* be exposed to every stimulus pair.

The use of the method of constant stimuli in the measurement of similarity among stimuli poses a special problem with regard to the interpretation of the judgment of two objectively identical stimuli. The results of this study indicate that the modal percentage of "same" judgments to objectively identical stimuli of the Morse code is less than 100 and that pairs of objectively identical stimuli differ significantly as far as the frequency with which these pairs elicit "same" judgments is concerned. The manner in which Estes and Burke (2), and Bush and Mosteller (1) conceptualize stimulation is useful in this respect. These writers assume that each stimulus is composed of a number of elements. At each stimulation *S* samples from that population of elements. When the stimulus is viewed in this fashion, the number of "same" judgments made to a pair of stimuli may be thought of as related to the number of stimulus elements which the two stimulus samples have in common. For objectively "different" stimuli, the number of elements common to the two *stimulus samples* is directly related to the number of elements which are common to the two *stimuli*. For objectively "identical" stimuli, the number of "same" responses may be thought of as a positive function of the total number of elements in the stimulus. This is to be expected because the probability of identical elements in repeated sampling is inversely related to the total number of elements from which one samples. The "same" responses to "identical" stimuli could therefore be considered as an inverse index of the psychological complexity of the stimulus. The finding of a correlation of about

— .69 between the relative frequency of such judgments and the number of substitution errors during paired-associate training argues for the utility of this index.

Summary

Similarity among aural signals of the International Morse Code was determined by use of the method of constant stimuli with two response categories (same: different). The *Ss* ($N = 598$), without prior code training, judged the 1,296 two-signal permutations which may be formed from the 36 common Morse signals. All signals were presented at high tone speed. The measure of stimulus similarity provided by this method was evaluated in terms of substitution error data from two studies in the literature of Morse code training and in terms of like data from two experiments performed at this laboratory. The data, for the levels of training which were attained, support the following conclusions:

1. The number of substitution errors made to a stimulus in a paired-associate situation is an increasing function of the average similarity of that stimulus to the other stimuli in the paired-associate list. This function for the greater part of the observed similarity range appears linear.

2. The number of substitution errors made to a stimulus is a decreasing function of the relative number of times that stimulus is judged "same" when presented twice in rapid succession.

The outcome of this experiment suggests that a measure of stimulus similarity based upon a psychophysical technique such as the method of constant stimuli may be useful for both applied and theoretical purposes.

References

1. Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58,** 413–423.
2. Estes, W. K., & Burke, C. J. A theory of stimulus variability. *Psychol. Rev.*, 1953, **60,** 276–286.
3. Gagne, R. M. The effect of sequence of presentation of similar items on the learning of paired associates. *J. exp. Psychol.*, 1950, **40,** 61–73.
4. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47,** 196–229.
5. Gibson, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30,** 185–200.
6. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
7. Keller, F. S. Studies in International Morse Code. 1. A new method of teaching code reception. *J. appl. Psychol.*, 1943, **27,** 407–415.
8. Keller, F. S., & Taubman, R. E. Studies in International Morse Code. 2. Errors made in code reception. *J. appl. Psychol.*, 1943, **27,** 504–509.
9. Plotkin, L. Stimulus generalization in Morse code learning. *Arch. Psychol.*, 1943, **40,** No. 287.
10. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44,** 430–444.
11. Yum, K. S. An experimental test of the law of assimilation. *J. exp. Psychol.*, 1931, **14,** 68–82.

(Received December 27, 1955)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

THE ROLE OF CONTEXT STIMULI IN VERBAL LEARNING[1]

LLOYD R. PETERSON AND MARGARET JEAN PETERSON

Indiana University

A discrimination experiment involves stimulus components other than those designated by the experimenter as the relevant cues. The function of these extraneous stimuli is of significance for theories of discrimination learning. Cognitive theories maintain that aspects of the stimulus situation to which the organism does not have a set to respond on certain trials are not established as cues during those trials (**4**). On the other hand S-R reinforcement theories hold that any discriminable component of the situation that is differentially reinforced will acquire cue value if the organism's receptor orientation is such that he is stimulated (**2, 6**).

Several verbal-learning experiments offer evidence bearing on this issue. It has been found that a stimulus not mentioned in the instructions, but correlated with a particular response, may acquire cue value for that response (**1, 5, 7**). Thus, in a recent study, Weiss and Margolius instructed their *Ss* to respond with English words to nonsense syllables appearing on distinctively colored cards in a paired-associate learning task (**7**). Retention tests leave no doubt that the colors aided recall and relearning, but it is noteworthy that relearning the verbal materials without the colored backgrounds resulted in no savings over original acquisition. The superiority of incidental cues over instructional cues implies that the *Ss* may have instructed themselves to respond to color rather than the nonsense syllables. This methodological difficulty holds for all studies using this design. An unknown proportion of the *Ss* is permitted to substitute another task for the instructional task. A cognitive theorist can hold that it has not been demonstrated that noninstructional stimuli can acquire discriminative status at the same time that instructional stimuli are being established as cues.

The present experiment was designed to insure learning of the instructional task, while still permitting a test for cue value of stimuli incidental to this task. Four nonsense syllables were used as instructional stimuli to be associated with four verbal responses. The nonsense syllables appeared on two colors of paper, the colors being context stimuli. Each color was paired consistently with two nonsense syllables and thus with two responses. The acquisition task could not be learned by associating context stimuli with the responses. S-R reinforcement theories would predict that each context stimulus should become a cue for a class of two verbal responses. This cue value could be measured by noting responses to the colors presented apart from the acquisition nonsense syllables. In order to test a quantititive prediction from one specific S-R reinforcement theory, one of the two responses following a particular color occurred twice as frequently as the other during acqui-

[1] This study was facilitated by a research grant from the Graduate School of Indiana University.

sition trials. A statistical model would then predict that probability of responding at the asymptote of learning should approach the probability of reinforcement (2).

Method

Subjects.—The *Ss* were 33 female volunteers from introductory psychology classes at Indiana University. Data are reported for the 32 subjects who met the criterion in acquisition.

Apparatus.—Learning materials were presented on a memory drum mounted behind a 24-in. by 36-in. black screen. A sliding shutter with ¾-in. by ¾-in. aperture permitted exposure of six different lists without changing the paper roll.

Learning materials.—Table 1 lists the learning materials. The nonsense syllables were of 20% association value or less (3), and were selected to minimize the possibility of responding to a single letter. Nonsense syllables in acquisition had no letters in common with those used in the transfer test. None of the nonsense syllables had letters in common with any of the nouns. Nonsense syllables were typed in upper-case letters on colored paper the size of the aperture. The responses were typed in lower-case letters on white paper.

Acquisition.—The *Ss* were instructed not to let the colored backgrounds distract them. The sign-up sheets had carried the title "Distraction Experiment." The *Ss* were told to concentrate on learning to speak the word which would follow a nonsense syllable before the word appeared. Each 2-sec. appearance of a nonsense syllable was followed by a 2-sec. appearance of the response alone. Nine pairs were presented in each revolution of the drum, an interval of 4 sec. elapsing between each block of nine. Serial order was randomized. One of the two responses associated with a given color occurred twice as frequently as the other response in each double block of 18 trials. Learning was carried to a criterion of correct anticipation of one double block.

Association test.—After a rest interval of approximately 50 sec. each color was presented alone until a response was given. The *Ss* were instructed to respond with the first of the English words they had just learned that occurred to them as they looked at a dot on colored paper. The interval between trials was 20 sec. Order of presentation was balanced.

Transfer test.—After an interval of 50 sec. there followed a new learning situation in which *Ss* were again instructed to learn to anticipate the word that would follow a nonsense syllable. The same responses used in acquisition were now paired with new nonsense syllables. For two of the four pairs the nonsense syllables appeared on a color which had been previously appropriate for the responses which followed. With the other two nonsense syllables, two new colors, yellow and brown, were used. Within each of the classes, appropriate (A) color and inappropriate (I) color, one of the two responses had been previously a high-frequency response (H) and the other a low-frequency (L) response. Five randomized lists were presented each *S*, a given paired associate appearing twice in every list, though not consecutively. The interval between lists was 8 sec.

TABLE 1

Learning Materials

	Stimuli	Color	Responses
Acquisition	VIH	red	foot (L)*
	VEJ	red	salt (H)*
	XEH	green	lamp (L)*
	XIJ	green	moon (H)*
Transfer	QUB	brown	foot
	QUW	red	salt
	ZUB	green	lamp
	ZUW	yellow	moon

* H refers to the stimulus-response pairs presented twice as often as those designated by L.

Results and Discussion

In the association test 22 of the 32 *Ss* gave a response to the color red which had been paired with red in acquisition. An equal number gave an appropriate response to green. Test of observed-versus-chance response level in either case gives a chi square of 4.4, which is significant at the .05 level. The summary of data in Table 2 shows that it was the high-frequency response of acquisition which was generally given to the color. The low-frequency response together with the inappropriate responses occurred less frequently than the chance level of .25. Tests of observed-versus-chance level for high-frequency responses yield values of chi square of 18.1 and 10.6 for red

TABLE 2

PROPORTIONS OF RESPONSES TO EACH COLOR IN THE ASSOCIATION TEST

Color	Responses			
	Salt (H)	Foot (L)	Moon (H)	Lamp (L)
Red	.53	.16	.16	.16
Green	.16	.16	.47	.22

and green, respectively, establishing significance at the .01 level for either case.

The results of the transfer test listed in Table 3 corroborate the findings in the association test. The analysis of variance summarized in Table 4 indicates that both color and frequency were significant factors. Individual *t* tests show the high-frequency response to an appropriate color to be significantly different at the .01 level from the other three responses. Differences among the latter are not significant.

The results support an S-R reinforcement interpretation of discrimination learning. Partially irrelevant context stimuli were established as cues at the same time that the instructional stimuli were acquiring discriminative status. The prediction from the statistical model that the asymptotic response level would approach the probability of reinforcement was not confirmed. However, the high level of error in the association test suggests that an asymptote may not have been reached. The assumption upon which the prediction was contingent was then not fulfilled.

TABLE 3

MEAN NUMBER OF CORRECT ANTICIPATIONS ON TRANSFER TEST

Color	Frequency			
	High		Low	
	Mean	SD	Mean	SD
Appropriate	6.87	1.84	5.06	2.12
Inappropriate	5.56	2.56	4.81	2.53

TABLE 4

ANALYSIS OF VARIANCE OF CORRECT ANTICIPATIONS ON TRANSFER TEST

Source	*df*	*MS*	*F*
Frequency (F)	1	51.257	16.215**
Color (C)	1	18.757	4.168*
Subjects (S)	31	10.633	
F × C	1	8.509	3.076
F × S	31	3.161	
C × S	31	4.500	
F × C × S	31	2.766	
Total	127		

* Significant at the .05 level.
** Significant at the .001 level.

Although the establishment of cue value in any degree for the incidental stimuli is evidence against cognitive theory, it remains for the S-R reinforcement theorist to derive the faster learning in the case of the instructional stimuli from previously established principles of his theory. There would seem to be no possibility of differential receptor orientation for the two sets of stimuli, since the color appeared around and within the letters. Neither is it reasonable that nonsense syllables are more easily discriminated than colors. A possible explanation is that the instructions result in implicit rehearsal behavior which increases the number of pairings of instructional stimuli with responses. One test of this explanation would involve the use of nonverbal stimuli as the instructional cues to minimize rehearsal behavior.

SUMMARY

The cue value of a context stimulus not perfectly correlated with any single response during paired-associate learning was investigated. The *Ss* were instructed to learn verbal responses to four nonsense syllables. Two background colors were the context stimuli, each color being paired with two responses. Each color was found to have acquired discriminative status for the more frequently occurring response of the two associated with

it. The results are interpreted as favoring a S-R reinforcement theory of discrimination learning.

References

1. Dulsky, S. G. The effect of a change of background on recall and relearning. *J. exp. Psychol.*, 1935, **18,** 725–740.
2. Estes, W. K., & Burke, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50,** 81–88.
3. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–267.
4. Lashley, K. S. An examination of the 'continuity theory' as applied to discrimination learning. *J. gen. Psychol.*, 1942, **26,** 241–265.
5. Pan, S. Influence of context upon learning and recall. *J. exp. Psychol.*, 1926, **9,** 468–491.
6. Spence, K. W. Continuous versus noncontinuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **47,** 271–288.
7. Weiss, W., & Margolius, G. The effect of context stimuli on learning and retention. *J. exp. Psychol.*, 1954, **48,** 318–322.

(Received December 30, 1955)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

STRENGTH OF SECONDARY REINFORCEMENT AS A DETERMINER OF THE EFFECTS OF DURATION OF GOAL RESPONSE ON LEARNING[1]

DAVID R. POWELL, JR.[2] AND CHARLES C. PERKINS, JR.

Kent State University

Spence's acute analysis of the role of secondary reinforcement (the reinforcing properties acquired by stimuli which precede reinforcement) in delayed reward learning (**9**) seems so promising that it appears worth while to extend this approach to other problems. The experiments reported below are tests of such an extension to the effects of duration of the goal response and/or magnitude of reward on learning.

In his analysis of delayed reward learning Spence assumed that (*a*) all reinforcement is immediate, and (*b*) secondary reinforcement is acquired not only by the stimulus followed immediately by reinforcement, but also by similar stimuli as a result of generalization of secondary reinforcing properties. That is, secondary reinforcement shows stimulus generalization just as does response strength.

An understanding of the implication of such an approach may be facilitated by considering Fig. 1

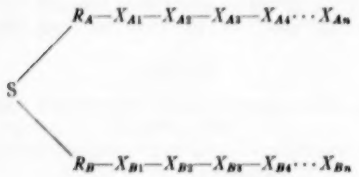


FIG. 1. Generalized schema for alternative responses to a single stimulus.

where R_A and R_B are two alternate responses which may occur to stimulus situation S. X_{A1}, X_{A2}···X_{An} are the sequence of stimulus situations which follow R_A; X_{B1}, X_{B2}···X_{Bn} represent the sequence of stimuli which follow R_B. This figure may be used to stand for a number of different experimental conditions. For example, it represents a delayed-reward study if X_{A4} has greater reinforcing properties than X_{B4} and if X_{A1}, X_{A2}, X_{A3}, X_{B1}, X_{B2}, and X_{B3} all are "neutral" or have equal reinforcing properties. In the typical delayed-reward experiment, X_{A4} would stand for a goal box containing food and X_{B4} an empty end box. X_{A1}, X_{A2}, X_{A3}, X_{B1}, X_{B2}, and X_{B3} could each represent 10-sec. periods during which S is in a delay box.

Under these delayed-reward conditions R_A will be strengthened at the expense of R_B, even though reward is not immediate, through one or both of two processes. (*a*) Acquisition of secondary reinforcing properties by X_{A3} which generalize more to X_{A1} than to X_{B1}. Of course this will occur only if X_{A1} is more similar to X_{A3} than is X_{B1}. Evidence that this process will strengthen R_A has been obtained by N. P. Perkins (**8**) and Grice and Goldman (**4**). (*b*) The acquisition of secondary reinforcing properties by X_{A3} which will in turn be acquired by X_{A2}, etc. R_A will then be strengthened at the expense of R_B after the secondary reinforcing properties have been passed back through the sequence to X_{A1}. The shape of learning curves obtained with

[1] This study was conducted as the first author's master's thesis. The second author is primarily responsible for preparing this report.

[2] Now with the American Greetings Corporation.

different spatial delays (6) and temporal delays (7) supports this interpretation.

Let us now consider how duration of the goal response (e.g., eating) resembles delayed reward.[3] The stimulus situation during successive 10-sec. intervals may be represented by X_{A1}, X_{A2}, etc. Then if R_A is followed by a long goal response (e.g., 150-sec. eating) and R_B is followed by a short goal response (e.g., 10-sec. eating), X_{A1}, $X_{A2} \cdots X_{A12}$, and X_{B1} represent presence of food and the accompanying eating while X_{B2}, X_{B3}, etc. represent "neutral" stimuli.

It will be noted that in a magnitude of reward study with a long goal response following R_A and a short one following R_B just as in a delayed reward study X_{A1} and X_{B1} have equal initial reinforcing properties but that subsequent events following X_{A1} are more reinforcing than subsequent events following X_{B1}. Thus if S receives long eating periods in one goal box and short eating periods in a distinctively different goal box, sight of the first goal box should come to have greater reinforcing properties.

Experiment I

The above interpretation implies that any stimulus situation which has preceded a long eating time will be more reinforcing than a stimulus situation which has preceded a short eating period and thus that a response followed by the former stimulus will be learned in preference to a response followed by the latter. Experiment I is designed to test this deduction when the quantity of food and duration of eating during test trials are the same following either response.

Method

Apparatus.—A starting box, straight runway, single choice-point E maze, and interchangeable goal boxes were used. The runway, maze alleys, and starting box were painted a medium gray. The runway was 35.5 in. long.

The maze was E shaped with the starting box on the middle arm and the goal boxes at the end of the "upper" and "lower" arms of the E. The length of the starting arm of the maze was 24 in.; the outer arms were 7 in. between turns and delay boxes. The section joining the maze arms was 47.5 in. over all. All alleys (runway and maze) were 4.5 in. wide and 3.75 in. deep. The black goal box was 10 in. wide and 12 in. long. Its floor level was approximately .5 in. higher than that of the alleys. The white goal box was 4 in. by 13 in., and its floor level was about .5 in. lower than that of the alleys. A fine mesh wire screen covered the floor of the white goal box to add a distinctive tactual stimulus. The entire apparatus was covered with hardware cloth.

The food trays were glass coasters resting on thin slabs of wood. One end of the slab was built up and squared to form a block so that when the food tray was withdrawn, this block formed a portion of the back wall of the goal box. Both food trays were the same color as the corresponding goal boxes.

Strings attached to the guillotine doors and food trays permitted their control by E from the recording table 4 ft. behind the starting box.

There were three sources of light in the experimental room. A 7-w. bulb in a "Snapit Nite-Lite" shaded socket (manufactured by Cable Electric Products, Inc., N. Y.) was placed 10.5 in. directly above each food tray. The bulb over each goal box was so shaded that light from it did not strike the interior of the other goal box. The third light, a 7.5-w. bulb on the recording table, gave no direct light and little indirect light on the part of the experimental room in which the apparatus was placed.

The apparatus used during preliminary training was the starting box, the straight alley,

[3] It is assumed that magnitude of reward effects result from past learning based on different durations of the goal response following large and small reward. This assumption follows from our analysis and is supported by a minor unpublished study conducted by R. Rehula in the Kent State laboratory. He ran three rats in a single choice-point maze with a large pellet which was removed after 20 sec. in one arm. In the other arm S obtained a small pellet which was supplemented by several additional pellets delivered one at a time as soon as the preceding one had been consumed. In this arm S was permitted a total of 120-sec. eating. Although Ss showed a predominance of turns to the large pellet side in the early stages of training, each S later developed a 100% preference for the other (small pellet, 120-sec. eating) side.

and the black and white goal boxes. The same starting box and goal boxes were used with the straight runway and maze. The location of the goal boxes was the same throughout the experiment.

The maze was centered in the experimental room. The white goal box was always in a position that required *S* to make two right turns to approach it; the route to the black box involved two left turns. The white goal box reflected a greater amount of light in the arm of the maze adjacent to it than did the black goal box. This difference was apparent from the choice point to normal human vision. However, the results to be reported indicate that this cue did not affect responses on the first maze trial.

Subjects.—The *Ss* were six male and six female, experimentally naïve, albino rats from the colony maintained by the Kent State University Department of Psychology. All *Ss* were approximately 5 mo. old when training was initiated.

Preliminary training.—The male *Ss* were reduced to approximately 75% of their normal body weight by small daily feedings. The lighter females were reduced to 80%. About 15 days were required to reduce *Ss* to the weights at which they were kept throughout the experiment by daily feedings immediately after the experimental period.

The preliminary training consisted of 40 trials in the straight alley over 11 days. Half the *Ss* received 20 trials ending with a long (150 sec.) goal response in the black box and 20 trials ending with a short (10 sec.) goal response in the white goal box. The opposite pairing of goal boxes and eating times (150 sec. W and 10 sec. B) was used for the remaining *Ss*. There were 15–30 min. between successive trials during preliminary training.

The food tray was present and filled with mash (Purina Lab Chow mixed with water) every time *S* entered the goal box. It was withdrawn at the end of 150 or 10 sec., depending on the assigned duration of the goal response for *S* on that trial. Throughout the experiment *S* was removed from the goal box within 5–10 sec. after the food tray had been withdrawn. The order of presentation of the goal boxes for any given day was irregular to prevent *S* from responding differentially to the sequence of trials.

Training.—Four trials were given on each of maze-training Days 1–5. *Every maze trial* was terminated with a 20-sec. eating period *in either goal box.* After *S* had made two responses to the same side, it was forced in the opposite direction to assure two entrances to each goal box every day. There were approximately 20 min. between successive trials.

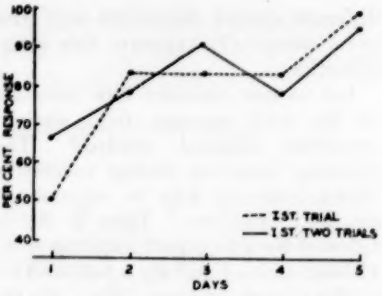


FIG. 2. Results on training series of Exp. I. The ordinate indicates the percentage of responses to the side with the goal box in which longer eating time had been provided during preliminary training.

Results

Figure 2 presents the results of the training series. Each point on the broken line represents the percentage of *Ss* that made their first run to the goal box which had been associated with the goal response of greater duration during preliminary training. Each point on the solid line represents the percentage of responses to this side on the first two runs of the day.

Examination of Fig. 2 reveals that on the first trial of the first training day in the maze, 50% of the *Ss* ran to either side. Thus, there is no indication that transfer from preliminary training produced an initial turning preference dependent upon spatial orientation, or the difference in the amount of light reflected to the section of the maze visible from the choice point.

Further examination shows that on Day 2, 10 of the 12 *Ss* made their first run to the goal box which had previously been assigned the goal response of greater duration. A 10:2 split would occur only 158 times in 4096 due to chance alone ($P < .04$). This is evidence of exceptionally rapid learning of a turning response. That it was not due to the eating time in

the goal box during maze-training trials is apparent since each *S* received two runs to each goal box during this training and *all* of these trials terminated with a 20-sec. goal response. It should be emphasized that there was rapid learning even though turns to either goal box resulted in a reward.

On the first two trials of the last three days combined every *S* ran above chance (50%) to the goal box which had been associated with the longer eating time in preliminary training.

The results are consistent with those of D'Amato (1) and Lawson (5) which demonstrated that goal boxes containing large amounts of food acquire greater secondary reinforcing properties than goal boxes containing small amounts of food if the same *S*s are run to both boxes. The results of these studies in which the goal boxes contained no food when their secondary reinforcing properties were measured were not as clear cut as the present results perhaps because of the more rapid extinction of differential secondary reinforcing properties of the goal boxes.

Experiment II

If entrance into Goal Box A is regularly followed by a long eating period (e.g., 150 sec.) and entrance to Goal Box B is followed by a short eating period (e.g., 10 sec.), the magnitude of the difference in reinforcing properties of entrance to A and B after a given number of trials will be a function of the distinctiveness of the difference between the two boxes. At one extreme there would be no generalization of secondary reinforcing properties from A to B or vice versa so that the reinforcing properties acquired by entrance to one (A) would be just the same as if there had been no entrances to the other (B). However, at the opposite extreme, if there were complete generalization from one goal box to the other the reinforcing properties of entrance to either would be the same as if all entrances (with long and short eating periods) had been to this same goal box.

The same factors are involved if entrance to a certain goal box is followed by 150-sec. eating whenever R_A preceded entrance to the goal box and by 10-sec. eating whenever R_B preceded entrance to the goal box. If the stimulus aftereffects of R_A and R_B are not discriminably different after 10 sec. in the goal box, the reinforcing effect of goal-box entrance following R_A will depend on the duration of eating on *all* previous entrances to the goal box. Thus, in this case, if R_A and R_B have preceded goal-box entrance an equal number of times, entrance to the goal box will have the same reinforcing effect on R_A if 10-sec. eating followed R_A and 150-sec. eating followed R_B as if 150-sec. eating followed R_A and 10-sec. eating followed R_B.

Experiment II was designed to test this deduction. By preliminary training in which long and short eating times each followed 50% of the entrances to either box, and by interspersing maze trials with straight-alley runs so that on all trials approximately the same eating time per goal-box entrance had followed entrance into either goal box, the secondary reinforcing properties of entrance to each goal box were roughly equated. It should be noted that the procedure does not eliminate the possibility that *S*s might discriminate between entrance to the goal box on maze trials and on straight-alley trials. Thus, there remains some chance of differential secondary

reinforcing properties of entrance to the different goal boxes when used with the maze.

Method

Apparatus.—The apparatus was the same as that used in Exp. I.

Subjects.—Twelve male, experimentally naïve, albino rats from the Kent State colony were used as *Ss*. They were approximately 5 mo. old when training was initiated.

Experimental procedure.—Table 1 summarizes the experimental procedure. The table does not indicate the within-day sequence; it merely shows *S*'s total daily training. The color of the goal box on maze-training days was as shown in the table for only half the *Ss*; the longer eating time on maze trials and on the runway was in the opposite goal box for the other *Ss* to counterbalance for possible color preferences.

On the 20 maze-training days two trials in the maze were irregularly interspersed with two trials in the straight alley. On each of these days half of the *Ss* were assigned a goal response of 150-sec. duration in the black goal box when used with the maze, and a 10-sec. goal response when the black goal box was used with the straight alley. The same *Ss* were assigned 10 sec. in which to eat in the white goal box when it was used with the maze and 150 sec. in which to eat when this goal box was used with the straight alley. The other six *Ss* were given the opposite pairings of eating time and condition (runway or maze) for each goal box color to counterbalance for possible color preference, etc. The purpose of interspersing straight-alley trials was to maintain the same average eating time for all previous entrances to each of the goal boxes. The general procedure for this experiment was the same as that for the previous one except as specifically indicated.

TABLE 1

THE PROCEDURE FOR EXP. II

Days	Trials	Apparatus	Goal Response
P 1	1	Straight alley	W 150 sec.
	1	Straight alley	B 150 sec.
P 2	1	Straight alley	B 10 sec.
	1	Straight alley	W 10 sec.
P 3–6	1	Straight alley	W 150 sec.
	1	Straight alley	W 10 sec.
	1	Straight alley	B 150 sec.
	1	Straight alley	B 10 sec.
M 1–20	1	Straight alley	B 150 sec.
	1	E maze	B 10 sec.
	1	Straight alley	W 10 sec.
	1	E maze	W 150 sec.

Note.—B and W represent the black and white goal boxes, respectively. P represents preliminary training; M represents maze training. The schedule represents that for one-half the *Ss*; the other half had a reversal of the black and white goal-box relations during training.

Results

The results of the maze training are presented in Fig. 3. Each point in the figure represents the percentage of *Ss* turning to the goal box in which they were permitted to eat for 150 sec. on the first maze trial each day. Only 9 of the 12 *Ss* chose the side on which they received the longer eating time on more than half of their free runs on Days 1–20 or 16–20 of maze training ($P > .1$). The change in performance from Days 1–5 to Days 16–20 was even less reliable. The absence of any reliable evidence of a preference for the turn followed by a longer eating time is in line with the hypothesis that strength of reinforcement is affected by the duration of the goal response only indirectly through altered reinforcing properties of sight of the goal object.

On the last day of training in Exp. I, 11 of the 12 *Ss* made their first two runs to the side previously associated with the longer eating time; at the same stage in training (Days 9 and 10) just half of the *Ss* in Exp. II made both free runs to the arm with the longer eating time. This difference between the results of the two experiments is significant ($P < .035$, one-tail test) according to Fisher's exact test (**2,** p. 96). Thus, our results indicate that if differential secondary reinforcement is eliminated, duration of the goal response does not have the marked effect which it has if differential secondary reinforcement is operative.

It should also be noted that the slight trend toward learning of the

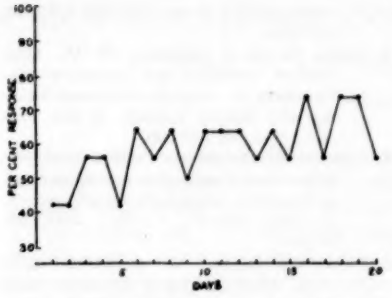


FIG. 3. Results of training series of Exp. II. Each point represents the percentage of *Ss* making their free runs to the side on which they were given the longer eating time on maze runs.

response followed by long eating in Exp. II could be the result of the gradual development of a discrimination between maze and runway trials. With this type of procedure—which does not eliminate the possibility of such a discrimination—the occurrence of significant learning with a larger *N* or longer training would not contradict (or support) the present theoretical interpretation.

DISCUSSION

The results of both experiments support the hypothesis that the effects of magnitude of the goal object on performance result indirectly from the duration of the goal response. Furthermore, they suggest that duration of the goal response as such does not affect the degree to which preceding responses are strengthened, but rather that duration of the goal response on preceding trials determines the secondary reinforcing properties of sight of the goal object (in a specific context). This in turn affects the extent to which antecedent responses are strengthened.

These findings are in harmony with the results of Greene's study of magnitude of reward and acquisition of black-white discrimination (**3**). His *Ss* were given 30 preliminary short-runway training trials in which half the rats received a large pellet in a black box and a small pellet in a white box. The other *Ss* received a large pellet in the white box and a small one in the black. All *Ss* were then given discrimination learning trials with black positive. Half the *Ss* with each preliminary training condition received a large pellet as reward and half received a small pellet. Greene obtained a significantly higher percentage of correct responses for those *Ss* which had received the large pellet in black during preliminary training. The size of the pellet used as reward during discrimination training had no apparent effect on performance.

The present results also suggest the utility of a clear distinction between magnitude (and quality) of reward and strength of reinforcement, the former defined operationally in terms of the nature of the goal object and the latter in terms of its effect on strengthening (or weakening) preceding responses. Under certain specific conditions, such as use of the same size food pellet throughout preliminary training and the main training series, there appears to be a close correspondence between magnitude of reward and strength of reinforcement. However, no such correspondence is obtained under conditions such as those of the present experiment.

SUMMARY

Two experiments were conducted to test implications of the assumption that duration of the goal response affects performance only by producing differential secondary reinforcing properties of sight of the goal object and associated stimuli. In Exp. I twelve rats were given preliminary training in a straight runway. On half of these trials they entered one goal box and were permitted to eat for 150 sec.; on other preliminary runs they entered a distinctively different goal box containing the same amount of food and were permitted to eat for 10 sec. In subsequent training in a single choice-point maze all *Ss* quickly developed a preference for the goal box previously associated with the longer eating time even though entrance to either goal box was followed by a 20-sec. eating time.

In Exp. II, 12 rats were given preliminary straight-runway training in which entrance into each goal box was followed by a 150-sec. eating

time on half the runs and a 10-sec. eating time on the other half. In the subsequent maze runs, one turn led to one goal box and a 150-sec. eating period; the opposite turn led to the other goal box and a 10-sec. eating period. Along with this training, straight-runway trials were given so that approximately the same average eating time had always followed entrance to either goal box throughout training. No significant preference for either turn developed.

It was concluded that duration of the goal response determines the reinforcing p operties of sight of the goal object and that strength of reinforcement depends on the reinforcing properties of sight of the goal object and is independent of the duration of the goal response on that particular trial.

References

1. D'Amato, M. R. Secondary reinforcement and magnitude of primary reinforcement. *J. comp. physiol. Psychol.*, 1955, **48,** 378–380.
2. Fisher, R. A. *Statistical methods for research workers.* Edinburgh: Oliver and Boyd, 1946.
3. Greene, J. E. Magnitude of reward and acquisition of a black-white discrimination habit. *J. exp. Psychol.*, 1953, **46,** 113–119.
4. Grice, G. R., & Goldman, H. M. Generalized extinction and secondary reinforcement in visual discrimination learning with delayed reward. *J. exp. Psychol.*, 1955, **50,** 197–200.
5. Lawson, R. Brightness discrimination performance and secondary reward strength as a function of primary reward amount. *J. comp. physiol. Psychol.*, 1957, **50,** 35–39.
6. Perkins, C. C., Jr. An experimental study of the effect of delay of reward on maze learning. Unpublished master's thesis, State Univer. of Iowa, 1942.
7. Perkins, C. C., Jr. The relation of secondary reward to gradients of reinforcement. *J. exp. Psychol.*, 1947, **37,** 377–392.
8. Perkins, N. P. A study of the role of secondary reinforcement in maze learning. Unpublished master's thesis, State Univer. of Iowa, 1943.
9. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54,** 1–8.

(Received January 9, 1956)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

MUSCULAR TENSION AND THE HUMAN BLINK RATE

DONALD C. KING AND KENNETH M. MICHELS

Purdue University

This study is concerned with motor or response facilitation. Previous studies have suggested two findings which Meyer (**4**, p. 204) calls the "two molar laws of interaction." The first of these laws is that interaction varies directly with the magnitude of the inducing response. This principle is illustrated by the findings of Courts (**1**), who reported that when his *S*s squeezed a hand dynamometer with varying force, knee jerk amplitude increased as the force of gripping increased. The second principle suggested is that interaction varies directly with the proximity of the simultaneous responses. A study by Freeman (**3**) illustrates this principle. He found that facilitation of finger oscillation was greatest if tension was induced in the adjacent fingers.

Meyer (**4**) postulated that the blink rate could be used as an index of muscular tension. He concluded that the interaction of responses depends upon the convergence of simultaneous patterns of neural impulses upon the motor pathways of the central nervous system. The motor channel for the eyeblink responses is bordered on one side by the motor structures for the face and on the other side by those for the hand (**5**). Thus, if Meyer's conclusions concerning response interaction are correct, facial or hand activity should increase one's rate of blinking, at least until a very high level of activity is reached.

Meyer states that directly induced muscular tension should result in an increase in the blink rates, but says that so far there are no data regarding this postulate (**4**, p. 215). The present study investigated the question of whether directly induced muscular tension affects the human blink rate.

METHOD

The number of eyeblinks of each *S* was recorded during six separate 1-min. periods. The experiment was conducted in two parts. In Part 1, *S*s' blink rates were recorded for 1 min. during two different conditions of gripping two hand dynamometers simultaneously and during a rest period. The *S*s were unaware that their blink rates were being recorded.

In Part 2, the data recorded, order and type of performance, and length of recording was the same as in Part 1. However, *S*s were aware that their blink rates were being recorded and were told to attempt to inhibit their blinking during all three experimental conditions.

Subjects.—The *S*s were 66 students in a beginning college psychology course. Of these 25 were women, 41 men. Their median age was 19 yr.

Apparatus.—The *S* was seated before a table on which rested two hand dynamometers (not fastened to the table) positioned so that their grips extended over the edge of the table. Two wires, fastened to each dynamometer, led to an adjoining room in which *E* sat. Both rooms were windowless. The *S*'s room was lighted, *E*'s was not. The *E* could see *S* clearly through a one-way screen, but *S* could not see *E*. Directly in *S*'s line of vision on the one-way screen between the two rooms was a 2-in. white square on which *S* focused during the experimental trials. In *E*'s room was a constant speed chronograph on which *E* recorded each blink, and a timer used to start and stop each trial.

Procedure.—Upon being seated *S* squeezed both dynamometers simultaneously as hard as possible with palms up while keeping the main body of the dynamometers resting on the table. The *E* then computed ¼ and ½ of *S*'s maximum grip for each hand. The *S* was then read the following instructions for Part 1.

"This is a test to see how accurately and consistently you can grip hand dynamometers. You will work with two different settings on the dynamometers during this part of the experi-

ment. You will attempt to maintain a constant pressure on the dynamometers. If your grip becomes too weak or too strong, the needle will hit one of the stoppers with an audible click. If this happens, increase or decrease your grip as is appropriate. Try to keep the needle between the markers at all times.

"You will work under each of the conditions for 1 min. and will have a formal 1-min. rest period. During the experimental trials and the formal rest period keep your eyes on the white square in front of you. This is in order to keep conditions constant for all subjects. Do not talk while you are squeezing the dynamometers or during your formal rest period.

"I will say 'ready,' 'set,' 'go,' and 'stop.' When I say 'ready,' place your hands on the dynamometers. When I say 'set,' position both needles between the stoppers while still looking down. When I say 'go,' look up at the white square and attempt to maintain your strength of grip so that neither needle touches one of the stoppers. When I say 'stop,' relax and wait for the next condition. . . . Are there any questions?"

No recording of accuracy of grip was taken. The *E* recorded a blink on the chronograph only if *S*'s eyes closed fully. A 3-min. period was maintained between adjacent experimental conditions.

At the conclusion of Part 1 *S* was read the following instructions for Part 2.

"During this last part of the experiment, I am also going to record your blink rate. You will go through the same sequence that you just finished. During both conditions of gripping and your formal 1-min. rest period, try to inhibit your blinking as much as possible. Again keep your eyes focused on the white square during the time you are gripping the dynamometer and during your formal 1-min. rest period. Keep your eyes open normally. Do not squint. Are there any questions?"

Experimental design.—A replicated 3 × 3 Latin square design was used in this experiment. The *S*s were assigned to the rows of the design; the columns defined sequence; and the treatments were assigned to the Latin letters of the square. For a given *S* the sequence of treatments was the same in Part 1 and Part 2.

Results

Results of the analyses of variance of the data obtained for Parts 1 and 2, respectively, are shown in Table 1. The analyses were made separately since the estimates of within-*S* residual variance in Parts 1 and 2 differ significantly. Note that the only significant values of *F* are for the treatment effects.

The values of the over-all means for both parts of the experiment are presented in Fig. 1. The two profiles of means differ significantly.

Tests for significance of differences between means were made by the

TABLE 1

Analyses of Variance

Source	*df*	Part 1		Part 2	
		MS	*F*	*MS*	*F*
Between *S*s	(65)				
Rows	2	184.384	—	4.641	—
Replications	21	366.407		51.150	
Rows × Reps.	42	345.008		51.890	
Within *S*s	(132)				
Columns	2	24.490	1.387	7.944	—
Columns × Reps.**	42	17.662	—	12.511	—
Treatments	2	119.460	5.492*	119.460	15.093*
Treatments × Reps.**	42	21.751	—	7.915	—
Residual**	44	22.874		11.015	
Total	(197)				

* Significant at .01 level.

** For Part 1, test for homogeneity: $F_{max.} = \frac{s^2_{max.}}{s^2_{min.}} = \frac{22.874}{17.662} = 1.295$; 3, 42 *df*; not significant. Pooled error term mean square = 20.782, *df* = 128. For Part 2, $F_{max.} = 1.581$; 3, 42 *df*; not significant. Pooled error term mean square = 10.488, *df* = 128.

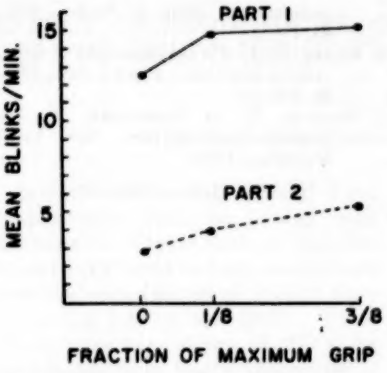


FIG. 1. Mean number of blinks for each condition of tension in Part 1 and Part 2.

Duncan procedure (2). In Part 1 the number of blinks while relaxed is significantly less than that while squeezing the dynamometers, but the difference between the two levels of tension is not significant. In Part 2 each of the three means differs significantly from the other two. The .05 level was used in making all tests by this procedure.

The 1-min. response periods for each *S* in Part 1 were broken into 20-sec. periods to see if *S*'s responses tended to change with time for any of the experimental conditions. No significant change in response was found for any condition.

In an attempt to determine to what extent *S*'s ability to refrain from blinking was dependent upon his rate with no attempt at blink inhibition, the total number of blinks for each *S* on Part 1 was correlated with his total number of blinks on Part 2. The correlation was .50. The value is significantly different from zero at the .01 level.

DISCUSSION

A significant rise in the blink rate while at rest with the induction of muscular tension was found in two different experimental settings. The utility of the eyeblink as a group measure or index of generalized muscular tension is enhanced by the results of Part 2. Although the total number of responses was less for each of the three experimental conditions, as compared to Part 1, the relationships were not only maintained but were pronounced to the point that all conditions were significantly different from each other. It appears that *S* need not be unaware that his blink rate is being recorded.

The findings obtained are consistent with Meyer's theory of response interaction. It seems unnecessary to regard the motor cortex of the brain as anything more than a closed system, following the laws of spatial summation, to explain the results.

While the present experiment clearly demonstrates the relationship between muscular tension and blink rate, it should be noted that this relationship is meaningful only for groups of *S*s. The large between-*S*s variance indicates that extensive individual differences exist and that the blink rate may not be a particularly good index for estimating the degree of tension in any given *S*, e.g., measuring individual *S* changes in muscular tension as an estimate of fatigue in industry.

SUMMARY

The blink rate of 66 college students was recorded for 1 min. during three experimental periods; while *S*s were relaxed, and while *S*s squeezed two hand dynamometers simultaneously with ⅛ maximum grip and then with ⅜ maximum grip. These recordings were taken first while *S*s were unaware (Part 1) and then while they were aware (Part 2) that their eyeblinks were being recorded. In Part 2, *S*s attempted to inhibit their eyeblinks.

During Parts 1 and 2 the *S*s blinked significantly more often during both tension conditions than during the relaxed condition. In Part 1 the rate of blinking was not significantly different for the two tension conditions, but during Part 2 the blink rate was significantly greater for the greater condition of induced tension. The correlation between blink-rate scores for Parts 1 and 2 was .50.

References

1. Courts, F. A. The knee-jerk as a measure of muscular tension. *J. exp. Psychol.*, 1939, **24**, 520–529.
2. Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.
3. Freeman, G. L. The facilitative and inhibitory effects of muscular tension upon performance. *Amer. J. Psychol.*, 1933, **45**, 17–52.
4. Meyer, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.
5. Penfield, W., & Rasmussen, T. *The cerebral cortex of man.* New York: Macmillan, 1950.

(Received January 16, 1956)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

EFFECT OF FOOD, DRINK, AND TOBACCO DEPRIVATION ON THE CONDITIONING OF THE EYEBLINK RESPONSE

C. M. FRANKS

University of London Institute of Psychiatry, Maudsley Hospital

Several studies (**12, 13, 14**) have demonstrated that *Ss* having high scores on the Taylor Scale of Manifest Anxiety (**15**) tend to form conditioned eyeblink responses more readily than *Ss* scoring low on this scale. These results have been shown (**3, 11**) to be predictable from Hull's theory (**9**) if it is assumed that scores on Taylor's anxiety scale are indicative of manifest anxiety and therefore of drive strength (*D*). If this Hullian prediction is correct then it should hold for variations in *D* other than anxiety, since in any specific situation the value of *D* is assumed to be determined by both relevant needs and the total aggregate strength of existing irrelevant needs.

One way of altering *S*'s degree of hunger and thirst drives is by depriving him of food and drink for a long period. As Atkinson (**1**, p. 54) points out, should scales of hunger, sex, thirst, and other drives be developed, then the scores could be combined to give a more complete indication of the total drive strength of *S*.

Hull (**10**) has developed the theory that the motivational variables *V* (stimulus intensity), *K* (incentive level), and *D* interact with each other and with *H* multiplicatively to determine *E* and hence response strength. If this is correct then any changes produced in *D* should produce corresponding changes in response strengths whatever the values of *V* and *K*. Spence (**11**, p. 13 ff), however, has rightly criticized this multiplicative hypothesis as being based upon insufficient evidence Instead, he has suggested that *V*, *K*, and *D* summate additively to determine response strength. If this is correct then any predictions concerning the effects of varying *D* will have to take into account the values of *V* and *K*. Only if *V* and *K* are low can a variation in *D* appreciably alter response strength.

In the present study *D* was increased by depriving *Ss* of food, drink, and tobacco for approximately 24 hr., keeping *V* and *K* as low as possible (this precaution is of course unnecessary if Hull's multiplicative hypothesis is accepted). The measure of response strength used was the frequency of conditioned eyeblink responses. The prediction, using the Hullian model in its original form or in the manner suggested by Spence, is that deprived *Ss* will produce more CR's than *Ss* who have been eating, drinking, and smoking normally. This makes the not unreasonable assumption that hunger and thirst do not introduce other unknown factors which may vitiate conditionability.

METHOD

Subjects.—The *Ss* were 60 undergraduate, male, university students, ranging in age from 18 to 26. All were naïve with regard to psychological theory and none was a student in a department of psychology.

Conditioning apparatus.—The apparatus has been described in detail elsewhere (**4**). It is sufficient to state here that the UCS consisted of an air puff, lasting 500 msec., delivered at a pressure of 65 mm. of mercury from a 2.5-mm. internal diameter polythene tube situated approximately 2 cm. from the right eye. The CS was a pure tone, delivered to both ears through

a pair of padded high-quality earphones at a frequency of 1100 cps for 800 msec. at an intensity of 65 db. above *S*'s hearing threshold. The air puff was so arranged that it began 350 msec. after the commencement of the tone. These time intervals were accurately controlled by means of an electronic timer. The air was supplied from a cylinder and puffed into the eye by means of an electronically operated gas valve. The polythene tube ran from the gas valve to a small aperture in the right lens of a pair of plain glass spectacles worn by *S*. The eyelid movements were recorded by means of a photoelectric cell attached to the same lens as the polythene tube (**7**). The recording milliammeter was so arranged that eyelid movements and the occurrences of the CS and the UCS all appeared in the same record.

Procedure.—All 60 *S*s had been instructed to go without any food or drink at breakfast. Upon arrival at the Institute of Psychiatry at approximately 9 A.M., each *S* was randomly assigned to either a "normal drive" or a "high drive" group. This assigning was not done by *E*, who remained in complete ignorance as to which group each *S* belonged. All *S*s were submitted to an easy, nonfatiguing, and nonintensive series of tests during the day and then brought to the present *E* in the early evening. It must be stressed that this was the first contact *E* had with *S*s that day and that all *S*s had been carefully instructed not to give him any indications as to whether they had or had not been eating, drinking, and smoking during the day. The "normal drive" group consisted of 32 *S*s; these *S*s followed the normal British meal pattern of midmorning coffee, hot cooked lunch, and afternoon tea and biscuits, so that they could not be considered as being hungry or thirsty when brought to *E* in the early evening. The 22 members of this group who were smokers were told to smoke throughout the day in their usual manner. The "high drive" group consisted of 28 *S*s, the high-drive state being effected by total deprivation of food, drink, and tobacco throughout the day, although everyone else around them was eating, drinking, and smoking in the usual manner (17 members of this group were normally smokers). Thus by 6 P.M. these 28 *S*s had received no sustenance since their dinner the previous evening. It can be safely assumed that they were both hungry and thirsty by the time they reached *E* and that as a group they were hungrier and thirstier than the "normal drive" group.

The *S*s were brought to *E*'s room about 6 P.M. and were paid an agreed fee, traveling expenses, and a fixed-meal allowance, irrespective of whether they had been eating or not. The *E* told them that they would shortly be eating, but that first they were to be taken into a quiet room and measures taken of how they were "able to relax under various conditions." (Since both tone and air puff were of only moderate intensity, and were neither threatening nor unpleasant, and since all *S*s were paid student volunteers who, as a group, had already met *E*, it seems reasonable to assume that both *V* and *K* were maintained at a minimal level throughout the experiment.)

Each *S* was then taken to the soundproof conditioning laboratory (**5**) and seated comfortably in an armchair at a small table so arranged that his field of vision was largely confined to the plain walls of a booth constructed around this table. His head rested against an adjustable head rest and his feet were supported by a padded footstool. The laboratory could be divided into two parts by means of a sliding curtain. In one part was *S*, in the other was *E* and most of the apparatus. The conditioning procedure has been described more fully elsewhere (**6**). First *S*'s hearing threshold was ascertained, then brief tests for initial eyelid sensitivity to sound and for pseudoconditioning were carried out. Partial conditioning was used, the sequence being such that the reinforcement ratio was approximately 60% throughout the reinforcement trials. Thirty reinforcement trials were given, interspersed by 18 test trials (called acquisition trials) consisting of the CS alone. The intertrial interval varied from 20 to 30 sec. with a mean of approximately 25 sec. After the 30 reinforcement trials had been given, *S*s were given a further series of 10 consecutive test trials (called extinction trials). The complete session lasted approximately 30 min.

Results and Discussion

Four *S*s had to be rejected, two because they blinked excessively, making the records unscorable; one because he gave only irregular UCR's,

TABLE 1

Number of Conditioned Eyeblink Responses

Group	*N*	Acquisition		Extinction	
		Mean	*SD*	Mean	*SD*
High Drive	25	6.96	3.44	2.79	2.43
Normal Drive	31	7.35	4.97	2.81	2.76
t		.281		.014	

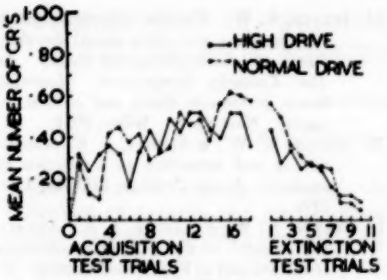


FIG. 1. The mean number of CR's given on each test trial.

and one because of an experimental error during testing. Two of the four *Ss* were smokers, one in each group. Table 1 presents the mean number of CR's for the remaining *Ss* to the 18 acquisition CS's and the 10 extinction CS's. For neither acquisition nor extinction measures are the mean number of CR's significantly different; on the contrary, they are remarkably similar as the *t* ratios indicate. Even the directions of the differences are opposite to that predicted according to the Spence-Hull formulation.

Figure 1 presents the conditioning curves of acquisition and extinction for both groups. The similarity in conditioned-response behavior is also readily apparent from an inspection of the two curves. It would seem reasonable to conclude that attempts to increase *D* by deprivation of food, drink, and tobacco fail to produce any increase in the conditioned eyeblink response of normal *Ss*.

These results, like those of Buchwald and Yamaguchi (**2**), who investigated the effects of water deprivation upon habit reversal in the rat, fail to give any support to the Hull-Spence theory of conditioning. They are, however, consistent with an alternative theory which relates conditionability not to drive strength, but to introversion-extraversion. It has already been demonstrated (**6**) that both eyeblink and PGR conditioning are related *not* to neuroticism but to the degree of introversion-extraversion present, as measured by the scales of Guilford (**8**).

Both the Hull-Spence and the introversion-extraversion theories predict that anxious *Ss* would condition better than nonanxious, but for different reasons: Spence and his colleagues (**3, 11, 12, 13, 14**) because of the greater drive strength of the anxious *Ss*, and the present writer because of the greater degree of introversion of these *Ss*. In the present experiment it is drive strength that is manipulated and not introversion-extraversion; hence, the results, as well as casting doubt on the Spence-Hull theory, provide further support for the introversion-extraversion theory of conditioning.

SUMMARY

The experiment was conducted to test the Spence-Hull prediction that increased drive strength would increase conditionability.

Sixty normal male undergraduates were allocated to one of two groups, a "high drive" group who were deprived of food, drink, and tobacco throughout the day, and a "normal drive" group who were allowed to eat, drink, and smoke in their usual manner. All *Ss* were then conditioned in a sound-proof laboratory, using the frequency of conditioned eyeblink responses to a tone stimulus as the measure of conditioning. The results failed to support the Spence-Hull theory of conditioning, but were in agreement with an alternative theory which relates conditionability to introversion-extraversion. Further experimental evidence was cited in support of this alternative theory.

REFERENCES

1. ATKINSON, J. W. Comment on Professor Farber's paper. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Nebraska Univer. Press, 1954.
2. BUCHWALD, A. M., & YAMAGUCHI, H. G. The effect of change in drive level on habit reversal. *J. exp. Psychol.*, 1955, **50**, 265–268.
3. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.

4. Franks, C. M. *An experimental study of conditioning as related to mental abnormality.* Unpublished doctor's dissertation, Univer. of London, 1954.
5. Franks, C. M. The establishment of a conditioning laboratory for the investigation of personality and cortical functioning. *Nature,* 1955, **175,** 984–985.
6. Franks, C. M. Conditioning and personality. *J. abnorm. soc. Psychol.,* 1956, **52,** 143–150.
7. Franks, C. M., & Withers, W. C. R. Photoelectric recording of eyelid movements. *Amer. J. Psychol.,* 1955, **68,** 467–471.
8. Guilford, J. P. *Inventory of factors STDCR.* Beverly Hills, Calif.: Sheridan Supply Co., 1940.
9. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
10. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1950.
11. Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *The Kentucky Symposium: Learning theory, personality theory and clinical research.* New York: Wiley, 1954.
12. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.,* 1953, **45,** 116–119.
13. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.,* 1951, **42,** 183–188.
14. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.,* 1951, **41,** 81–92.
15. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48,** 285–290.

(Received January 3, 1956)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

THE EFFECT ON PROBLEM SOLVING OF SUCCESS OR FAILURE AS A FUNCTION OF CUE SPECIFICITY [1]

RAMON J. RHINE

The RAND Corporation, Santa Monica, California

This research is concerned with the influence of certain cue and motivational factors on problem solving. Studies of complex intellectual processes (e.g., **1, 2, 7**) as well as considerable research on the effects of success and failure on learning, retention and perception (**3**) suggest that such influences are probable. If previous success or failure affect problem solving, it would be assumed that some residual or partial cue aspects of these conditions are present at a later time to activate an inhibitory or facilitative process. The degree to which these cues operate should depend upon how distinct they are; the more discriminable the stimuli, the more likely that they can cue associated responses.

Experiment I was primarily directed toward testing two hypotheses: (*a*) failure inhibits and success facilitates problem solving, and (*b*) the degree to which success and failure affect problem solving depends upon the similarity between the situation in which the cues for success and failure occur and the subsequent test situation. The first experiment also investigated possible effects associated with sex (**6**) or due to achievement motivation. For Exp. II it was hypothesized that with increased distinctiveness of cues, relatively little failure experience would be sufficient to decrease later achievement.

Experiment I

Method

Materials.—The problem materials were anagrams which were catagorized as easy, moderate, or difficult on the basis of the percentage of *Ss* solving each problem in preliminary tests. In order to test for cue effects, two classes of anagrams were included: those which could be rearranged to form the names of living creatures and those which could be rearranged to form the names of countries.

The measure of achievement motivation was the need-achievement test developed by the McClelland group (**4**). In addition, a short questionnaire dealing with affective reactions to the anagrams and with *Ss*' awareness of success or failure conditions was administered.

Subjects.—The *Ss*, 110 men and 74 women, were students drawn from an elementary psychology class, and assigned to one of five groups. Because there were available fewer women than men, one group contained no women.

Procedure.—All tests were administered in group sessions. Each group took the need-achievement test two days prior to the time when the anagrams were given to that group. Immediately after completion of the anagram test, all *Ss* filled out the questionnaire.

The last 20 anagrams (test anagrams), which followed the first 15 (pretest anagrams) without a break, were living creature problems of moderate difficulty and were identical for all five groups. However, the pretest anagrams differed for each group, either as to type or as to difficulty, as follows: Group DCr received difficult creature problems; Group DCo had difficult country problems; Group MCo, moderate countries; Group ECr, easy creatures; and

[1] The experiments reported here were part of a general program of research on thinking carried on under the direction of Dr. Donald W. Taylor in the Department of Psychology at Stanford University. These experiments were completed and reported (**5**) under Project NR 150-149, supported by Contract N6onr 25125 between Stanford and the Office of Naval Research. Permission is granted for reproduction, translation, publication, use and disposal of this article in whole or in part by or for the United States Government. The studies also were submitted as a dissertation in partial fulfillment of the requirements for the Ph.D. at Stanford University. The author is indebted to Dr. Donald W. Taylor for his aid with this research.

Group ECo, easy countries. Before the pretest, *Ss* were told that they could expect to solve about one-half of the problems. Of course, those who received difficult anagrams could not solve half, and those who received easy anagrams could solve more. The heading *living creature* or *country* preceded each anagram to indicate the class of solution possible.

Thus Groups DCr and DCo are failure groups, both having difficult pretest problems, but there is a greater similarity between the pretest and the test situation for Group DCr than for Group DCo. Similarly, Groups ECr and ECo, having easy problems, are success groups, with the pretest being more like the test situation for the former than for the latter group. Group MCo *Ss*, who were able to solve about as many problems as they were led to believe they would solve, make up the control group.

Results and Discussion

Details of the questionnaire were reported elsewhere (**5**). On questions dealing with affective reactions, Groups DCr and DCo did not differ nor did Groups MCo, ECr and ECo, but the former two gave less positive affective reactions to the anagrams than the latter three. Those questions concerned with *S*'s awareness of success or failure showed quite clearly that the level of difficulty intended for each set of problems was experienced as such.

The anagram score for *S* on the pretest series and also the test series was simply the number of anagrams solved. With sex as one variable and pretest conditions as a second, an analysis of variance was done with the test anagrams. Since Group ECo consisted entirely of men, it could not be used in this analysis. An F of 6.12 with 3 and 120 *df* indicates an effect due to pretest conditions, but there is no significant main effect due to sex ($P > .05$) and no reason to accept the hypothesis that success and failure interact with sex since the F for interaction was only 1.63 with 3 and 120 *df*.

A second analysis of variance was carried out in which *Ss* were classified along two dimensions: success-failure and class of anagrams. This analysis is summarized in Table 1. It is clear that the combined success groups do significantly better than the combined failure groups and that success-failure conditions interact with class of anagrams. Thus the success-failure effect depends upon the similarity between pretest and test situations since similarity was varied by using two different classes of anagrams.

TABLE 1

EXPERIMENT I: ANALYSIS OF VARIANCE FOR TEST ANAGRAM SCORES WITH SUCCESS-FAILURE CONDITIONS AND CLASS OF ANAGRAM AS VARIABLES

Source	*df*†	*MS*	*F*
Success-failure (S-F)	1	162.00	14.44**
Anagram class (A)	1	4.00	.36
S-F × A	1	50.50	4.48*
Within	124	11.26	

* $P < .05$.
** $P < .001$.
† For all double classification analyses, some *Ss* were randomly discarded in order to equalize the cell *N*'s.

Because the direction of possible mean differences is stated in the hypotheses, one-tailed t tests were used to compare all means. Group DCr differs from Group DCo beyond the .05 level and from all other groups beyond the .005 level. The only other difference was between Groups DCo and ECr at less than .05. That Group DCo did not differ from Group MCo is somewhat surprising. The pretest problems were very similar to the test problems in that both were anagrams, even though the class of anagrams differed. Apparently, at least with the degree of failure involved here, the failure must be associated with very specific cues in order to produce a significant effect. It is probable that less specific cues are also associated with failure and that the combined effect of a number of such cues would be

sufficient to produce an inhibiting influence if the degree of failure were larger.

Groups ECr and ECo were included to test the effect of success upon problem solving because it was thought that the easy pretest problems might produce more positive affect than that produced in the control group by the moderate problems. However, this condition was not realized as the questionnaire responses indicate. Therefore, the hypothesis that success facilitates problem solving can be neither confirmed nor rejected.

All need-achievement tests were scored by the writer using scoring system C (**4**, pp. 147–148). First, an interscorer reliability of .86 was established with a qualified scorer who had learned the system under McClelland.

One analysis of variance gave neither evidence to suggest that need-achievement differs from group to group nor any significant indication at the .05 level that need-achievement scores in the present sample interact with sex, although McClelland *et al.* (**4**, p. 178) report that "women get higher *n*-Achievement scores than men under neutral conditions." A second analysis gave no statistically significant indication that need-achievement as measured affects problem solving in interaction with success-failure conditions ($F = 1.36$ with 1 and 150 *df*), and no evidence that anagram solution varies as a function of need-achievement ($F = 2.29$ with 1 and 150 *df*). Thus need-achievement has not been shown to be related to problem solving either when taken alone or in combination with success-failure conditions.

TABLE 2

MEANS AND *SD*'S OF ANAGRAM SCORES: EXP. I AND II

Group	*N*	Pretest Mean	Test Mean	Test *SD*
Exp. I				
DCr	36	.81	8.28	3.23
DCo	35	1.43	9.71	3.66
MCo	35	7.80	10.29	3.18
ECo	38	13.24	10.79	3.23
ECr	40	12.38	11.30	3.08
Exp. II				
DCr′	36	1.22	8.81	3.09
DCr″	40	1.40	8.95	3.59
MCo′	34	8.44	11.00	3.85
DCrEA	32	.56* 6.03†	8.56	2.47
PC	31	—	10.48	3.38

* Difficult creature problems.
† Easy arithmetic-series problems.

EXPERIMENT II

The main purpose of Exp. II was to investigate the influence of increased distinctiveness of failure cues. In addition, a control was added for possible practice effects since it might be argued that the failure groups of Exp. I did less well because of less practice. Including this control was a conservative step because the argument for possible practice effects is weakened by the findings of Exp. I. Groups DCo, ECr, and ECo did not differ significantly from Group MCo, the control group, although Group MCo solved more pretest anagrams than Group DCo and fewer than Groups ECr and ECo.

Method

Materials.—The moderate and difficult creature anagrams which were used in Exp. I were also employed here. Seven easy arithmetic-series problems were also employed. Following the problems *Ss* were given a questionnaire similar to the one which was administered in the first experiment.

Subjects.—The *Ss*, 79 men and 94 women, were drawn from an elementary psychology class,

but none of them had participated in the previous study.

Procedure.—As in the first experiment, 20 moderate creature anagrams were given to five groups of *Ss*, four of these groups having first been given pretest anagrams. Groups DCr′ and MCo′ were replications of Groups DCr and MCo, respectively, of Exp. I. The conditions for Group DCr″ were like those for Group DCr (and Group DCr′) in that each received 15 difficult creature anagrams as a pretest. However, *Ss* of Group DCr″ were told that they could expect to solve between zero and three of the pretest anagrams. The aim of this instruction was to reduce to a minimum the experience of failure resulting from the fact that very few of the first 15 anagrams were solved.

Group DCrEA had eight difficult creature anagrams, alternated with seven easy arithmetic-series problems as a pretest. The contrast between easy arithmetic-series problems and difficult creature anagrams was intended to increase the distinctiveness of the failure cues. This group was told that it could expect to solve about one-half of each of the two kinds of problems.

Group PC, included as a control for practice effects, received as its test problems the same 20 moderate creature anagrams that all other groups had, but it was given no pretest anagrams whatsoever. It was told that solving 50% of the problems was average.

Results and Discussion

Responses to the questionnaire approximate the results obtained in Exp. I. However, even in the face of instructions intended to minimize failure, Group DCr″ contained a fairly high percentage of people (50%) who thought their performance on the pretest was rather poor.

A single-classification analysis of variance was carried out to test for differences between the means of the test anagrams. Table 3 summarizes these results. Since this analysis led to a significant difference, comparisons of all pairs of means were made by one-tailed *t* tests. These tests show that Groups DCr″, DCr′, and DCrEA, all of which had difficult pretest anagrams, differed from Groups MCo′ and PC at the .05 level or beyond.

TABLE 3

EXPERIMENT II: ANALYSIS OF VARIANCE FOR TEST ANAGRAM SCORES WITH PRETEST CONDITIONS AS THE VARIABLE OF CLASSIFICATION

Source	*df*	*MS*	*F*
Groups	4	40.99	3.57*
Within	168	11.48	

* $P < .01$.

Group DCr″ was included because it was thought that if feelings of failure could be kept to a minimum, the scores would be higher than those of Group DCr′. Such a difference, if obtained, would be evidence that a failure effect had occurred regardless of whether or not there was a practice effect. Since responses to the questionnaire indicate that Group DCr″ did seem to experience failure despite attempts to prevent it, this comparison could not be used to illuminate the question of practice.

Fortunately, there are other data bearing on this problem. That there was a real effect from failure, over and above possible practice effects, is shown by the fact that Groups DCr′ and DCrEA solve significantly fewer anagrams than Group PC.

Group DCrEA, like Groups DCr′ and DCr″, did poorer than Group MCo′ on the test problems. That is, when there are fewer failures, and these are brought into sharp relief, the failure effect still carries over. However, since Group DCr′ did not differ significantly from Group DCrEA, it can only be concluded either that a relatively small number of failure experiences will have an influence when made quite distinct or that a relatively small number alone will yield an effect upon problem solving. It is possible that failing most of 8 anagrams may have as much effect on later achievement

as failing most of 15 anagrams, even when the 8 are not alternated with problems of a different type. This would be true if there is a curvilinear relation within this general range between number of failures and the effect on later achievement. To the extent that such a relation is possible, but not probable, the former of the two alternative conclusions is favored.

Summary

Five groups of *Ss* participated in a first experiment which was designed to test the effect of motivational and cue factors on achievement with anagram problems. Two of these groups were failure groups that did poorly on pretest anagrams; two others were success groups that did well on pretest anagrams; and a fifth, the control *Ss*, did about as well as they expected on pretest anagrams. All five groups were given the same 20 test anagrams immediately after these experiences. One failure and one success group had pretest anagrams with solutions of the same class as the answers to the test problems, while the other failure and other success group had pretest anagrams of a different class.

Failure reduced later achievement when the cues associated with pretest and test anagrams were very similar. Neither sex nor achievement motivation influenced performance on the test problems.

In a second experiment five additional groups were given anagram problems, three groups under conditions which differed somewhat from those in the first study. With relatively few failures, there is still an inhibitory effect when the cues associated with failure are made quite distinct through contrast. It was also shown that it would be difficult to explain the failure effects on the basis of differential practice.

References

1. Cowen, E. L. Stress reduction and problem solving rigidity. *J. consult. Psychol.*, 1952, **16**, 425–428.
2. Lantz, B. Some dynamic aspects of success and failure. *Psychol. Monogr.*, 1945, **59**, No. 11 (Whole No. 271).
3. Lazarus, R. S., Deese, J., & Osler, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
4. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
5. Rhine, R. J. *The effect on problem solving of success or failure as a function of cue specificity.* Tech. Rep. No. 8 (Contract N6onr 25125). Stanford: Dept. Psychol., Stanford Univer., 1955.
6. Sweeney, E. J. *Sex differences in problem solving.* Tech. Rep. No. 1 (Contract N6onr 25125). Stanford: Dept. Psychol., Stanford Univer., 1953.
7. Thistlethwaite, D. Attitude and structure as factors in the distortion of reasoning. *J. abnorm. soc. Psychol.*, 1950, **45**, 442–458.

(Received January 9, 1956)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

MOTIVATIONAL PROPERTIES OF FRUSTRATION: III. RELATION OF FRUSTRATION EFFECT TO ANTEDATING GOAL FACTORS[1]

ABRAM AMSEL AND WILLIAM HANCOCK

Newcomb College, Tulane University

The term *frustration effect* (FE) refers to an observable increase in response vigor immediately following nonreward of a response which previously was rewarded. This increase in response intensity following frustrative nonreward has been attributed to a temporarily increased drive level due to the presence of an additional motivating condition, *frustration*. This is similar to a position taken by Skinner: "When we fail to reinforce a response that has previously been reinforced . . . we set up an emotional response—perhaps what is often meant by frustration" (**8**, p. 203). Two previous experiments in this series (**1**, **2**) and other experiments (**5**, **7**, **9**) support such a motivational interpretation of frustrative nonreward.

The assumption is made that only after some minimal number of rewards will nonreward of a response be frustrating. To account for this, frustration is conceptualized as being the result of an interaction between nonreward and a factor which has been developing during previous rewarded trials. Hull, in his last book (**3**) and earlier in an article with Livingston, Rouse, and Barker (**4**), identifies this factor when he refers to frustration as "the failure of an expectation from being realized." The "expectancy" construct in the Hullian system is the fractional antedating (anticipatory) goal reaction ($r_g \rightarrow s_g$). Presumably, then, not rewarding a response will elicit frustration to the degree that earlier rewards of that response have led to the conditioning of r_g to cues in an instrumental response sequence. If, during preliminary reward trials, conditions for development of r_g are favorable (and if conditions for its elicitation on test trials are favorable), nonreward should produce greater frustration than if conditions for r_g development and elicitation on test trials are less favorable. By definition, one set of conditions very favorable to development and subsequent elicitation of r_g in a runway situation exists when cues accompanying the instrumental response are identical to cues present during the consummatory activity which rewards that instrumental response.

The experiments reported here test the relevancy to the frustration effect of the similarity of runway (instrumental sequence) stimuli to goal stimuli. It is assumed that similarity of runway to goal stimuli affects the intensity of $r_g \rightarrow s_g$ evoked during an instrumental response sequence; hence, it would affect the magnitude of FE evoked with the nonreward of that instrumental response.

[1] This research was supported in part by a grant from the Tulane University Council on Research. The authors are indebted to Claire Ernhart and Charles Galbrecht for critical reading of the manuscript and helpful suggestions.

General Method

Apparatus.—A labeled floor plan of the apparatus used in Exp. I and II is shown in Fig. 1. This is a slight modification of the apparatus

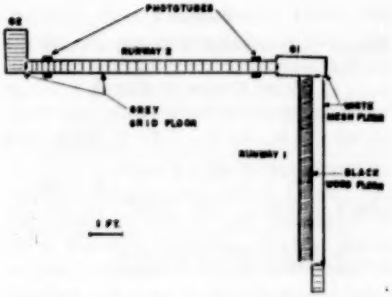


FIG. 1. Descriptive floor plan of apparatus.

employed by Amsel and Roussel (1). It consists, essentially, of two runways in series, the goal box of the first runway serving as the start box for the second. The first runway could be either white with a hardware-cloth mesh floor, to match stimulus qualities of G_1, or black with a smooth floor. It was 6 ft. long and 3 in. wide. The apparatus was 5 in. high throughout. Other dimensions in inches were: SB—12 × 3 × 5; G_1—18 × 6 × 5; Runway 2—96 × 3 × 5; G_2—15 × 6 × 5.

The floors of SB, Runway 2, and G_2 were constructed of ¼-in. brass rods spaced about ½ in. apart. This served to further differentiate them from Runway 1 and G_1.

Running time in Runway 2 was measured over a 7-ft. segment by placing photoelectric units 6 in. from G_1 and G_2. The phototubes were connected to a Standard Electric 1/100-sec. timer, and they were operated by the breaking of an infrared light beam, so their use involved no interfering visual stimulation.

A 40-w. fluorescent tube was suspended in line with Runway 1 and 3 ft. above it. This was the only illumination in the room.

Procedure.—The two experiments involve the same general procedure as an earlier experiment (1). Using an apparatus which is, essentially, two runway arrangements in series, where the goal box for the first runway is the start box for the second, the general procedure in both experiments was to measure FE in Runway 2 as a function of the stimulation in Runway 1 and reward or nonreward in G_1. The difference between the two experiments is in the consistency of reward in the first goal box (G_1). In Exp. I, 54 reinforcements in G_1 precede any nonreward in G_1, and the FE in Runway 2 with the introduction of nonreward in G_1 is related to the stimulus properties of Runway 1; in Exp. II, the Runway-1 response is rewarded 50% from the outset, and the magnitude of FE, which develops on the nonreward trials during partial reinforcement, is related to the stimulus properties of Runway 1.

Preliminary training was the same in both experiments. It consisted of 9 days of handling on the last 5 days of which *S* explored the apparatus for 5 min. each day. During this exploration period, the photoelectric system was operating so that *S*s could become adapted to the clicking noises of the starting and stopping of the clock. The deprivation schedule was in effect also during these 9 days.

The *S*s were run under thirst conditions defined as follows: After each day's trials *S*s were allowed to eat 10 gm. of food and drink water for 1 hr. About 21 hr. later, ½ hr. before the first of three trials on each day, *S* ate 4 gm. of dry food. This procedure appears to produce a highly stable thirst drive.

The incentive, two drops of water, was present on every trial in G_2 and was present in G_1 according to the experimental plan. The omission of reward in G_1 was the frustration-inducing operation. When *S* was not rewarded in G_1, *S* remained in G_1 for approximately the same length of time as on previous reward trial (about 6 to 9 sec., varying with the particular trial and *S*).

The *S*s were run in squads of about 20, with a minimum of 20 min. separating successive trials, three trials per day.

The *S*s for the two experiments were 84 male albino rats from a colony maintained by the Department of Psychology, Newcomb College, Tulane University. They were 3–4 mo. old at the start of the experiment. Of these, one *S* died leaving 83 *S*s from which data were collected.

STATISTICAL ANALYSES

During the last 54 trials of Exp. I, and through all trials of Exp. II, reward (R) and nonreward or frustration (F) trials were introduced according to the following arrangement: RFR, FRF, RRF, FRR, RFF, FRR. This constitutes one cycle of trials. For one-half the *S*s in each group the F's and R's were reversed, starting with FRF.

Within any cycle of R and F trials, there are three trials per day for six days. On any pair of days there are three R and three F trials, one R trial and one F trial appearing in each of the three trial positions (first, second, or third). In Exp. I there are nine such pairs of days (3 cycles) after the 18 consistent reward days; in Exp. II there are 18 such pairs of days (6 cycles).

For each pair of days for each *S*, a median of three F scores and three R scores was found. Then cycle medians for F and R were obtained for each *S* by taking the median of three suc-

cessive two-day medians. The curves plotted in Fig. 2 and 3 show means of such individual R and F cycle medians for successive cycles. At each point the individual *S*'s R score and F score is a median of three medians. This procedure yields stable values.

Two tests of significance were performed in each experiment. The first was a test for presence of the frustration effect within each group, B (black Runway 1) and W (white Runway 1). To do this, F and R scores were determined for each *S* over all trials in Exp. II or over all test trials in Exp. I. This was the mean of all cycle medians. Using the over-all F and R scores thus obtained for each *S*, a *t* for paired scores was performed within each of the groups of each experiment to test for the frustration effect, i.e., that F scores are significantly smaller than R scores.

The second test involved the difference (R − F) scores of the individual *S*s obtained in the first analysis. There was one such score for each *S* in each of the two conditions, and it represented the magnitude of the frustration effect for each *S*. To test the effect of antedating factors on the FE, a *t* test was performed on these difference scores between Group B and Group W in each of the two experiments.

Experiment I

Method.—The 42 *S*s of Exp. I were split into two groups. In Group W (white) the stimulus properties of Runway 1 and G_1 were identical except for spatial dimensions; in Group B (black) the stimulus properties of Runway 1 and G_1 were not similar, i.e., black alley against white goal box, smooth floor against hardware-cloth floor. All *S*s were run 54 trials (18 days, 3 trials per day) to consistent reward in G_1.

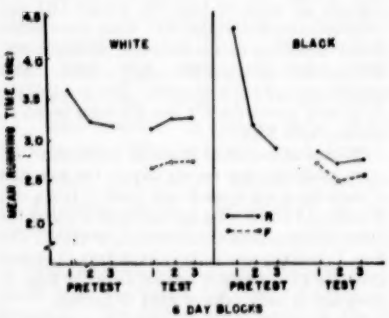


Fig. 2. Frustration effect on running time after 54 preliminary rewards for two conditions of Runway-1 stimulation (Exp. I).

TABLE 1

Means of Individual *S* Running Times and Differences in 1/100 Sec. on Three Test Cycles of Exp. I

Group	R	F	R − F	$W_{R-F} - B_{R-F}$
White	322	271	51*	32**
Black	274	255	19*	

* Significant at .01 level, one-tailed.
** Significant at .02 level, one-tailed.

Reward was always present in G_2. From Day 19 to Day 36, six cycles of R and F trials were run as previously described. Hence, the first experiment was a test of the effect of antedating goal factors on FE after a period of consistent reinforcement of the to-be-frustrated response.

Results.—The running times plotted in Fig. 2 are means of 21 individual *S* medians. It will be noted that the difference between the R and F curves is greater for Group W than for Group B, and that the difference, in either case, is present on the first block of test trials, and remains fairly constant through the three 6-day cycles of trials.

Table 1 represents means of individual *S* medians for R trials and F trials over all test trials for W and B conditions. It will be noted that under each condition, W or B, there is a highly significant frustration (R − F) effect (*t* values are 3.86 and 4.87, respectively). This substantiates the findings of Amsel and Roussel (1).

Of more particular pertinence to the purpose of this experiment is a comparison of the magnitude of the FE in the two groups. The difference in favor of Group W, for which, presumably, the antedating goal factors were strongest, is large (.32 sec.) and significant at close to the .01 level ($t = 2.32$, $df = 40$, $.02 > P > .01$ with one-tailed test). The validity of this last comparison might be called into question somewhat by

the fact that the running times for Group W were generally higher (in the pretest series and test series) than those of Group B. It could be argued that the difference in magnitude of FE between Groups W and B is an artifact, depending upon the generally inferior running speeds of Group W. An examination of the individual *S* data showed that the higher mean values in Group W could be attributed primarily to two *S*s who demonstrated extreme frustration effects. When the analysis was repeated without these two *S*s, reducing *N* to 19 for Group W, the Group W values of 322 for R, 271 for F, and 51 for R − F were reduced to 285, 252, and 33; but the *t* for R − F was raised from 3.86 to 7.02 and the *t* for $W_{R-F} - B_{R-F}$ was virtually unchanged (2.28).

Experiment II

Method.—The second experiment was identical to the first except for one factor relating to the first half of the trials (Days 1 to 18 or Trials 1 to 54). While, during these trials, *S*s of Exp. I were always rewarded in G_1, the *S*s of Exp. II were rewarded on 50% of the trials from the outset. Hence, *S*s of Exp. II—21 in Group W and 20 in Group B—ran six cycles of R and F trials, whereas *S*s of Exp. I had run three such cycles of trials from Days 19 to 36.

Accordingly, the purpose of this experiment was to study the *development* of FE in Runway 2 during a schedule of 50% reward in G_1 as a function of the presence of more or less effective stimuli for eliciting antedating goal responses in Runway 1. An earlier experiment by Roussel (6) had demonstrated that, although not initially present, FE develops when R and F trials are presented according to a 50%-reinforcement schedule from the outset of training in the Amsel-Roussel type of situation.

Results.—Figure 3 shows running time in Runway 2 after R and F in G_1 for Group W (strong antedating goal cues) and Group B (weak antedating goal cues) over the six cycles (108 trials) of the experiment. Each point is a mean of individual *S* medians as described earlier.

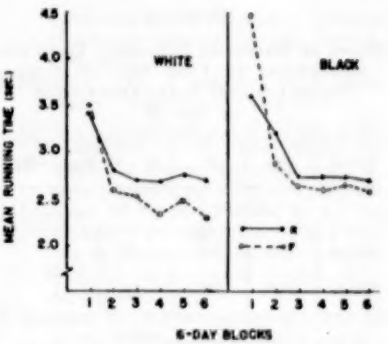


Fig. 3. Frustration effect on running time with no preliminary rewards for two conditions of Runway-1 stimulation (Exp. II).

Attention is called to the following: (*a*) No FE is apparent on the first cycle of trials (18 R and F trials); in fact, the F and R points are at almost the same level in Group W and are reversed, although not significantly, in Group B. (*b*) On the second cycle of 18 trials, FE is present and persists through the remaining cycles of the experiment. (*c*) The FE is apparently much greater in Group W than in Group B, and this larger FE in Group W reflects a greater reduction in running time following F in Group W as compared to Group B. The running times in Runway 2 following reward in G_1 were at almost identical levels through the experiment. This adds greatly to the validity of our interpretations of these data in terms of frustration effects.

Table 2 is a breakdown of the data into means and differences for the first three and the last three cycles. It is a further demonstration of the developing characteristic of FE under a partial reinforcement schedule. When the data of the first three cycles (27 R trials and 27 F trials for each *S*) are combined, neither Group W nor Group B shows a significant

TABLE 2

MEANS OF INDIVIDUAL S RUNNING TIMES AND DIFFERENCES IN 1/100 SEC. ON FIRST THREE AND LAST THREE TEST CYCLES OF EXP. II

Group	R	F	R − F	$W_{R-F} - B_{R-F}$
Cycles 1, 2, and 3				
White	297	288	9	20
Black	319	330	− 11	
Cycles 4, 5, and 6				
White	270	237	33*	22*
Black	273	262	11**	

* Significant at .01 level, one-tailed.
** Significant at .05 level, one-tailed.

FE, nor does $W_{R-F} - B_{R-F}$ approach significance. When the data of the last 3 cycles are analyzed, W_{R-F} is highly significant ($t = 4.64$), B_{R-F} is significant at about the .025 level with a one-tailed test ($t = 2.01$), and $W_{R-F} - B_{R-F}$ is significant at about the .01 level with a one-tailed test ($t = 2.37$).

DISCUSSION

The relevance of the variable defining r_g to the magnitude of FE was demonstrated under two conditions: first, where a large number of reinforcements of the to-be-frustrated response (running in Runway 1) preceded introduction of nonreward of that response; and second, where the Runway-1 response was rewarded and nonrewarded equally often from the beginning of training.

The major difference between these two cases was in the immediacy of appearance of FE with nonrewarded trials. In the first case, FE was evident immediately with introduction of nonreward; in the second case, nonreward was not apparently frustrating early in training but only after several trials of inconsistent reinforcement of the Runway-1 response.

The explanation for this difference between the Exp. I and II results seems clear. When nonreward of the Runway-1 response was introduced in Exp. I, r_g already was strongly developed on the basis of the 54 prior rewards of that response; therefore, in Exp. I, the earliest nonrewards in G_1 were frustrating and produced an invigoration of the Runway-2 behavior. In Exp. II, the first nonreward in G_1 had been preceded by, at most, one reward; consequently, the earliest nonrewards in G_1 were not frustrating—in the absence of sufficient previous rewards there was no prior development of r_g. After several rewards in G_1, under this condition of 50% reinforcement from the outset of the Runway-1 response, the cues in Runway 1 elicited r_g and nonrewards in G_1 evoked increasing frustration.

The results of Exp. I and II, and also the difference in results between Exp. I and II permit the tentative statement of the following hypothesis: Nonreward of an instrumental response is frustrating to the extent that stimuli accompanying the instrumental sequence leading to a goal elicit fractional antedating goal reactions appropriate to that goal.

One final point needs to be made in relation to the $r_g \rightarrow s_g$ construct. Although the fractional-antedating-goal-response construct has a certain theoretical elegance and validity—depending, as it does, on such well-established and basic phenomena as classical conditioning and stimulus generalization—and although it may even seem reasonable to suppose that implicit events corresponding to r_g actually occur, there has been little experimental work which links the construct to antecedent or to consequent manipulable-observable events. In the present experiments a variable is manipulated which defines the strength of r_G; but, what is more important, the strength of r_G, defined in terms of manipulated stimuli, is related to an instrumental response other than the one within which it is generated.

Summary

Using an experimental situation in which frustration effects due to nonreward have been demonstrated, and with a total of 84 male albino rats as subjects, two experiments were performed to test the relationship of strength of the frustration effect to strength of antedating goal factors.

The manipulated variable, which presumably affects the degree of elicitation of antedating goal responses, was similarity of cues in the instrumental response sequence (Runway 1) to cues in the goal (Goal 1) region. This variable was found to be relevant to the degree of the frustration effect (increased vigor in Runway 2) which followed nonreward of the first-runway response under two experimental conditions.

The results of the two experiments are interpreted as supporting an hypothesis that there is a positive relationship between the frustrating effect of nonreward and the strength of an antedating goal reaction preceding the nonreward.

References

1. Amsel, A., & Roussel, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.
2. Amsel, A., & Ward, J. Motivational properties of frustration: II. Frustration drive stimulus and frustration reduction in selective learning. *J. exp. Psychol.*, 1954, **48**, 37–47.
3. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
4. Hull, C. L., Livingston, J. R., Rouse, R. O., & Barker, A. N. True, sham, and esophageal feeding as reinforcement. *J. comp. physiol. Psychol.*, 1951, **44**, 236–245.
5. Lambert, W. W., & Solomon, R. L. Extinction of a running response as a function of distance of block point from the goal. *J. comp. physiol. Psychol.*, 1952, **45**, 269–279.
6. Roussel, J. Frustration effect as a function of repeated non-reinforcements and as a function of the consistency of reinforcement prior to the introduction of non-reinforcement. Unpublished master's thesis, Tulane Univer., 1952.
7. Sheffield, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.
8. Skinner, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, **57**, 193–216.
9. Stanley, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43**, 249–260.

(Received January 16, 1956)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

ANCHOR EFFECTS IN ABSOLUTE JUDGMENTS

CHARLES W. ERIKSEN[1] AND HAROLD W. HAKE[1]

The Johns Hopkins University

One of the most characteristic features of discrimination data obtained by the method of absolute judgments is the prominent anchor effects obtained for the stimuli on the ends of the stimulus continuum. Studies on such dimensions as size, brightness, and hue (**1**), loudness (**5**) and pitch (**9**) have shown that between five and nine stimulus categories can be absolutely discriminated without error along these dimensions. Yet on all of these dimensions, if the judgmental data are scaled on an equal discriminability scale (**6**), the majority of the discrimination is actually carried by a relatively few stimuli on each end of the continuum. It has been shown that the magnitude of these anchor effects is relatively independent of the number of stimulus and response categories (**2**).

Despite the prominence and the importance of these end effects, little is known concerning the factors that produce them. Garner (**4**) suggested that they might in part be due to a *response attenuation effect* that artifactually leads to unidirectional errors on the ends of the stimulus series. In the middle of the stimulus series, *S* can err in either direction in assigning a response to the stimulus while for the end stimuli *S* can err in only one direction. However, Garner (**4**) as well as Eriksen and Hake (**2**) have shown that this artifact is not sufficient to account for the observed anchor effects. The accuracy of discrimination for the end stimuli is much greater than can be accounted for on the basis of unidirectional distribution of error.

A second explanation for the anchor effects can be found in the behavior-theory model of discrimination and stimulus generalization. If it is assumed that for each value on the stimulus series a generalization gradient is set up in terms of the particular response that is assigned to that stimulus, there would be for the stimuli in the middle of the series a number of overlapping generalization gradients. Thus the occurrence of a stimulus in the middle of the series would tend to evoke a number of responses. Even if the correct response to that stimulus is assumed to have the greater habit strength, due to such factors as oscillation (**7**), competing responses would be evoked with a relatively high frequency. However, for the stimuli on the ends of the series, the generalization gradients would be asymmetrical with the result that there would be fewer competing responses and therefore fewer errors. Although there have been no systematic attempts to apply the stimulus-generalization model to absolute judgment data, it is apparent that this model would predict the end effects that are typically found.

There is a third explanation for the anchor effects that has been suggested by *S*s' accounts of their subjective behavior and by our own introspections while serving as *S*s in absolute-judgment tasks. This is the "subjective-standard hypothesis" which, simply stated, says that judgments are never made in a vacuum but are always made relative to a standard or reference level that is subjectively present.[2] The *S*, when

[1] Now at the University of Illinois.

[2] An explanation of anchor effects in terms of subjective standards or subjective frames of reference is as old as the problem itself. However, these explanations have seldom been explicitly formulated. The present purpose was to state, in a manner that was experimentally testable, how subjective standards are formed and how they contribute to judgmental accuracy.

confronted with a series of stimuli in an absolute-judgment task, selects a few stimuli in the series that he then uses as standards for judging the remaining stimuli. When a different stimulus is presented *S* attempts to recall one of these "standard" stimuli and uses it as a reference in judging the presented stimulus. Essentially he transforms the task into a comparative judgment using the recalled value of the selected standard as the comparison stimulus. If it is assumed that *S*s select the end stimuli to use as their subjective standards, the observed anchor effects are the expected outcome. The frequent attempt to recall the value of the end stimuli should increase the accuracy of recognition of these stimuli when they do occur. Also the recognition accuracy of the other stimuli in the series should be directly related to the distance of these stimuli from the end stimuli since the effectiveness of a standard is inversely related to the similarity of the standard and the comparison stimulus (**8**).

In the series of experiments reported below, data have been obtained which are definitely incompatible with the first two theories or explanations of anchor effects. The obtained data lend themselves most readily to an interpretation along the lines of the *subjective-standard hypothesis*.

Experiment I

Design

It would be possible to test empirically the adequacy of the three theories of anchor effects by means of an endless stimulus continuum. Such an endless continuum is available in the form of visible hues. By using a series of hues that vary from red through green, blue, and back to red again by varying through the purples, one can obtain a stimulus series that has no obvious breaks or end points.

Both the response attenuation and the stimulus-generalization hypotheses would predict that absolute judgments of such an endless continuum would be free of anchor effects. Since for any stimulus on this continuum *S* would be equally likely to err in either direction in the response he applies to the stimulus, there would be no response attenuation. Similarly, the stimulus-generalization gradients would be approximately symmetrical for each stimulus and thus response competition would be about the same for all stimuli.

The subjective-standard hypothesis would predict, however, that anchor effects would still be present. Subjects would still be expected to select a few stimuli or hues which they would attempt to recall with high frequency to use as standards in judging the other stimuli in the series. The particular hues selected might be expected to vary between *S*s, but due to cultural conventions and knowledge about primary colors, this selection would probably not be random.

Actually, it is possible to manipulate experimentally which of the hues *S*s are apt to select as their subjective standards by manipulating the response assigned to the specific hues. It is common practice to use the number series for responses in absolute-judgment studies since this has the advantage of providing *S*s with a familiar set of responses in which the sequential relations are already well known. The assumption in the present study is that in the absence of obvious stimulus values to select as subjective standards, *S*s would select their stimulus anchors in terms of the ends of the *number series* assigned them as responses. Thus, whatever hue is assigned the response "1" by *E* is most apt to be chosen as a subjective standard by *S*.

In the present experiment two groups of *S*s were used. Both groups made absolute judgments of the 20 Munsell hues shown in Table 1. Both groups also used the number series 1–20 to identify the 20 different hues, but the groups differed according to which response number was assigned to which stimulus. The *S*s in Group I were required to call the 5R stimulus by the response "1," the 10R stimulus by response "2," the 5YR stimulus by response "3," etc., and the number "20" was to be used for the

TABLE 1

MUNSELL SPECIFICATION OF THE 20 STIMULUS HUES, AND RESPONSE NUMBERS USED BY *Ss* TO IDENTIFY THEM*

Hues	Responses		Hues	Responses	
	Group I	Group II		Group I	Group II
5R	1	11	5BG	11	1
10R	2	12	10BG	12	2
5YR	3	13	5B	13	3
10YR	4	14	10B	14	4
5Y	5	15	5PB	15	5
10Y	6	16	10PB	16	6
5GY	7	17	5P	17	7
10GY	8	18	10P	18	8
5G	9	19	5RP	19	9
10G	10	20	10RP	20	10

* All hues had a brightness value of 5 and a saturation of 6 in the Munsell notation.

10RP. In Group II, *Ss* were instructed to call the 5BG by response "1" and 10BG by response "2," etc. Table 1 also shows the response numbers assigned to each stimulus by the two groups of *Ss*.

Method

Stimuli and procedure.—As can be seen from Table 1, the 20 hues used in the experimental series were so selected as to form a color circle and to thus provide a stimulus continuum which was endless. The stimuli were constructed by pasting 1-in. squares of Munsell papers in the center of 8 × 11-in. white cardboards. All hues were of equal brightness and saturation and the different hues were selected so as to be equally spaced on the Munsell system.

During the experimental sessions *S* was seated at a table and was screened from *E* by a partition. The stimuli were presented one at a time for *S*'s judgments, in an 8 × 8-in. aperture in the partition. The sole source of illumination was an overhead McBeth Daylight Lamp.

Prior to the experimental judgments *Ss* were given eight practice judgments on each of the stimuli. During these practice judgments the stimuli were presented in random order and *S* was informed as to the correct response after each of his judgments. Immediately following the practice session each *S* made 24 judgments of each of the 20 stimuli. During these experimental judgments *S* received no knowledge of the correctness of his judgments. The stimuli were presented in random order with the restriction that each of the 20 stimuli occur three times in each set of 60 judgments. Four duplicates of each of the stimulus hues were used in order to minimize the possibility that *Ss* might employ incidental cues, such as smudges on the stimulus cards, in making their judgments.

Subjects.—Subjects were male undergraduate students at The Johns Hopkins University enrolled in the elementary psychology course where they were required to serve a minimum number of hours as experimental *Ss*. A total of 10 *Ss* were used, 5 in each of the experimental groups.

Results and Discussion

The presence of anchor effects in the present judgmental data was determined by computing the accuracy with which each of the 20 hues was identified. In order to correct for differences in the frequency with which different response numbers were used, accuracy of judgment was computed as the percentage of times each response was used correctly. These percentages were computed for each *S* in the two groups. In Fig. 1 these percentages have been averaged by groups for each of the 20 hues.

As can be seen from this figure, both groups show anchor effects in their judgments. For Group I, judgmental accuracy is best for the 5R hue and for the 10RP hue with

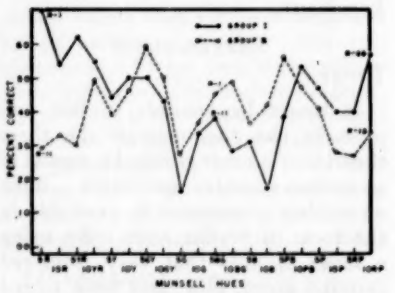


FIG. 1. Percentage of correct responses to each stimulus in the two experimental groups. These measures of discrimination accuracy were computed by dividing the number of times a response was used correctly by the total number of times the response occurred.

accuracy also being high for hues close to these values on the hue continuum. It is to be noted that these are the stimuli that *Ss* in this group were required to judge by the response numbers "1" and "20", respectively.

Anchor effects are not quite as prominent in the data from Group II, but here again the effects of responses upon accuracy of judgment is quite apparent if a comparison is made of the judgmental accuracy for hues 5R, 10RP, 10G, and 5BG.[3] Hues 5R and 10RP are much more accurately discriminated when they are called by responses "1" and "20" by *Ss* in Group I than when they are called by responses "10" and "11" by *Ss* in Group II. Similarly, the hues 10G and 5BG are more accurately judged by Group II *Ss* who label these hues by the responses "20" and "1," respectively, than by Group I *Ss* who apply the responses "10" and "11" to these hues. In general, the discrimination function for Group I tends to follow a U-shaped curve while the function for Group II suggests an inverted U function.

The significance of the above effects was tested by means of a modified three-way analysis of variance (groups, stimuli, and individuals). A summary of the results of this analysis is given in Table 2. In this analysis the subjective-standard hypothesis is tested by the interaction between groups and stimuli which, as is seen from the table, is significant beyond the .02 level. Since the interaction and its form was predicted, the obtained significance level is on the conservative side. Neither of the two experimental variables was significant by itself. There was no over-all difference between the two experimental groups, nor were the differences between stimuli significant when evaluated by the significant interaction term.

TABLE 2

SUMMARY OF THE ANALYSIS OF VARIANCE

Source	*df*	*MS*	*F*
Stimuli	19	.0695	.93
Groups	1	.0474	.95
Individuals	8	.0496	
Stimuli × Groups	19	.0743	1.95*
Pooled Individuals × Stimuli	152	.0381	
Total	199		

* $P < .02$.

The results of the present experiment quite clearly show that the discrimination accuracy obtained for a specific stimulus is in part at least a function of what response has arbitrarily been assigned to it. This finding is predicted from the subjective-standard hypothesis but is not accountable in terms of either the response attenuation or the stimulus-generalization hypothesis. If anchor effects in absolute-judgment data were solely due to either stimulus-generalization gradients or to response attenuation, there would have been no anchor effects in the present judgmental data. The subjective-standard hypothesis not only predicts the obtained anchor effects, but the location of the anchors are consistent with what would occur if *Ss* selected their subjective standards in terms of salient responses.

EXPERIMENT II

While the results of Exp. I are consistent with the subjective-standard hypothesis, the possibility exists that the response effects upon judgmental accuracy are due to response variables other than those deter-

[3] The effects of the experimental variables in determining anchor effects is somewhat obscured by what appears to be unequal discriminability among the stimulus hues. As can be seen from Fig. 1, certain hues are much better discriminated irrespective of their location with respect to the anchors. Thus, the hues in the range of 5RP and 10RP show good identifiability while the hue 5G is poorly discriminated by both groups.

mining *Ss*' choice of subjective standards. Garner, Hake, and Eriksen (3) have shown that not only can responses be operationally distinguished from perceptual processes, but also that care must be taken to distinguish between response and perceptual effects. Garner has suggested that the results of the above experiment might be due to what he terms response stabilization.[4] That is, responses on the ends of the response continuum may be more stable in that *S* uses them more carefully or sparingly. Alternatively stated, *S* may be better able to discriminate among his responses on the ends of the response continuum and thus be able to keep better track of how he has used them.

Volkman has made a somewhat similar suggestion of response effects.[5] He has suggested that in addition to stimulus generalization there is also response generalization along the response continuum. Thus, there would be asymmetrical response-generalization gradients for the end responses corresponding to the asymmetrical stimulus-generalization gradients for the end stimuli, and these response gradients would work in the same manner as has been discussed above for the stimulus gradients in improving discrimination for the end stimuli. This response generalization can be viewed as supplemental to the idea of response stabilization since it is essentially a theory to account for greater stability of the end responses. It was the purpose of Exp. II to determine whether these possible response effects contributed significantly to the anchor effects in absolute judgment.

[4] W. R. Garner. Personal communication. October 3, 1955.

[5] John Volkmann. Personal communication. November 11, 1955.

Method

The method and procedure were identical with those employed in Exp. I with the exception that in Exp. II the stimulus continuum consisted of a series of squares varying in size from 2 to 40 mm. square. There was a total of 20 squares, with the series stimuli increasing in size by steps of 2 mm. on each dimension. Ten additional *Ss* drawn from the same source were assigned five each to the two experimental groups. The *Ss* in Group I were instructed to call the smallest square in the series by the response number "1" and the largest square by the response "20" and to assign the intermediate numbers to the intermediate-sized squares in order of increasing size. In Group II, *Ss* began the number series in the middle of the stimulus series, assigning the number "1" to the square that *Ss* in Group I were to call "11" and, continuing in this manner, they called the largest square by the number "10" and the smallest by the number "11" with response "20" being assigned to the size Group I was calling "10."

Both the response stabilization and the response generalization hypothesis would predict that increased judgmental accuracy should be obtained for the stimuli identified by responses on the ends of the response continuum. Thus in Group II, increased judgmental accuracy should be obtained for the midsized squares in the series relative to the accuracy obtained for these squares in Group I. The subjective-standard hypothesis, on the other hand, would predict little or no effect of response assignment upon judgmental accuracy in the present experiment. It would be expected that with the opportunity available, *Ss* would select their standards in terms of the stimulus series where obvious ends are present, rather than from the response series. This is based upon the assumption that salient features present in the stimulus series are prepotent in determining *Ss*' choice of subjective standards.

Results and Discussion

In Fig. 2 is plotted the average percentage of times that a response was used correctly for each stimulus by each of the experimental groups. These percentages have been corrected for the total number of times a given response was used by dividing the number of times the response was used correctly by the total number of times the response occurred. As can

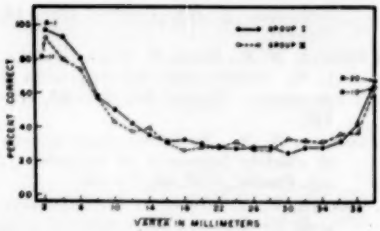


FIG. 2. Percentage of correct responses to each stimulus in the two experimental groups. These measures of discrimination accuracy were computed by dividing the number of times a response was used correctly by the total number of times the response occurred.

be seen, there is little or no difference in the judgmental accuracy for individual stimuli between the two groups. While the anchor effects are quite pronounced, they are essentially the same and occur for the same stimuli irrespective of the responses applied to these stimuli. The stimuli called by responses "1" and "20" by Group II are discriminated equally well by Group I *S*s who identify these same stimuli by responses "11" and "10."

This lack of effect of responses on the size judgment data is supported by a modified three-way analysis of variance. In this analysis neither of the experimental variables was significant nor did the interaction term approach acceptable levels of significance ($F = .97$, $P > .20$). Thus the results offer no support for the view that response variables, such as response generalization, contribute to any appreciable extent to anchor effects. In view of this, it does not appear very probable that such response variables can be used to account for the results of Exp. I.

Taken together the results of Exp. I and II are sufficient to show that the anchor effects in absolute judgments cannot be understood in terms of response attenuation, stimulus generalization, or even such response characteristics as response generalization or response stabilization. The findings from both experiments, however, are not only consistent with but are predicted by the subjective-standard hypothesis. This hypothesis emphasizes neither the stimulus nor the response factors in absolute judgments, but points out that either stimulus or response characteristics can be important in determining the individual *S*'s choice of stimuli to serve as his subjective standards. In the present studies it has been assumed that the ends of a stimulus continuum or, if they are absent, the ends of a response series, constitute salient perceptual characteristics that in turn determine *S*'s choice of standards.

In addition to describing the anchor-effect data, the subjective-standard hypothesis has several other advantages. The individual differences in discrimination accuracy under the absolute method have seemed somewhat too large to be attributed to differences in sensory acuity. By stressing *S*'s attempt to covertly recall the standards he has selected, this hypothesis not only relates the absolute-judgment task to processes important in learning, but it makes the magnitude of the individual differences more reasonable.

SUMMARY

One of the prominent features of discrimination data obtained under the method of absolute judgment is the anchor effects obtained on the ends of the stimulus continuum. Explanations for these anchor effects have been advanced in terms of response attenuation and stimulus generalization. In the present paper an alternative explanation, termed the subjective-standard hypothesis, is presented. This hypothesis states that *S*, when confronted with a series of stimuli in an absolute-judgment task, selects certain stimuli that he then uses as standards for judging the remaining stimuli. When a stimulus is presented, *S* attempts to recall one of these standard stimuli and uses it as a reference in judging the presented stimulus.

In Exp. I two groups of *S*s were required to judge absolutely a series of hues that formed a circular continuum. The responses consisted

of the number series 1–20. The groups differed in what response numbers were assigned to which hues. The results showed that anchor effects were obtained for this circular continuum and that the discriminability of a hue depended in part upon the response number that was assigned to it. These results are interpreted as being at variance with the response-attenuation hypothesis and the stimulus-generalization hypothesis, but were predicted by the subjective-standard hypothesis. The results of Exp. II successfully ruled out an explanation of the results of Exp. I in terms of such response factors as response stabilization and response generalization.

References

1. Eriksen, C. W., & Hake, H. W. Multidimensional stimulus differences and accuracy of discrimination. *J. exp. Psychol.*, 1955, **50**, 153–160.
2. Eriksen, C. W., & Hake, H. W. Absolute judgments as a function of stimulus range and number of stimulus and response categories. *J. exp. Psychol.*, 1955, **49**, 323–332.
3. Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.
4. Garner, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46**, 373–380.
5. Garner, W. R. An equal discriminability scale for loudness judgments. *J. exp. Psychol.*, 1952, **43**, 232–238.
6. Garner, W. R., & Hake H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
7. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
8. Hunt, W. A. Anchoring effects in judgment. *Amer. J. Psychol.*, 1941, **54**, 395–403.
9. Pollack, I. The information of elementary auditory displays. *J. acoust. Soc. Amer.*, 1952, **24**, 745–749.

(Received January 16, 1956)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

STIMULUS GENERALIZATION AFTER EQUAL TRAINING ON TWO STIMULI [1] [2]

HARRY I. KALISH [3] AND NORMAN GUTTMAN

Duke University

The present experiment is an attempt to investigate stimulus generalization gradients resulting from the reinforcement of two stimuli on the same continuum. The general problem derives from Hull's treatment of the summation of generalization gradients (**4, 5;** Postulates 4 and 5, Major Corollary I). These postulates state that overlapping generalization gradients will summate exponentially to yield a resultant gradient of greater response strength. Wolfle (**7**) has made a similar assumption to the effect that gradients generated from training at several points on the stimulus continuum will reinforce each other.

Until a recent study by Bilodeau, Brown, and Meryman (**1**), no explicit investigation of this problem had been undertaken. Their experiment used an instructed voluntary response rather than a conditioned response, presumably because of the technical difficulties inherent in the latter. The conditioning technique employed by the present authors (**2**), however, provides a more direct approach to the question.

METHOD

Subjects.—The *S*s were 22 experimentally naïve pigeons maintained by restricted feeding to 80% of their body weight under ad libitum feeding.

Apparatus.—The apparatus, a modified version of the Skinner automatic key-pecking apparatus, was the same as that used in a previous investigation by the authors (**2**).

Procedure.—The training and testing procedure of this experiment was essentially the same as that previously employed (**2**) with one major modification: each *S* was trained to peck at two monochromatic stimuli rather than one. Three groups of *S*s were used. One group ($N = 7$) was trained to 540 and 550 Mμ; the second group ($N = 8$) was trained to 530 and 560 Mμ; and the third ($N = 7$) to 520 and 570 Mμ.

Initial training was under continuous reinforcement. A total of 120 reinforcements (60 for each stimulus) were given during two training sessions, and the stimuli were alternated every 10 reinforcements. Following continuous reinforcement, 10 daily 30-min. periods of variable interval (VI) reinforcement (mean inter-reinforcement interval approximately one min.) were given,[4] alternating between the two stimuli every 5 min. Each period of VI training was divided into 1-min. intervals by 10-sec. blackouts. The change of stimuli was accomplished during the blackouts.

The test for generalization was carried out under extinction. Prior to the test for generalization, each stimulus was presented for three 30-sec. intervals during which two responses were reinforced. During the generalization test, each *S* in the three groups was presented with 10 test stimuli ranging from 500 to 590 Mμ in steps of 10 Mμ, in addition to a stimulus of 545 Mμ, midway between the two CSs for all groups. These 11 test stimuli were randomized within a series, and 12 different random series were presented to each *S*. Each stimulus presentation was 30 sec., and was followed by a 10-sec. blackout.

Following this test, *S*s were retrained under VI for three daily periods and on the fourth day were given a second generalization test identical to the first.

[1] This study was supported by grant MH-1002 from the National Institute of Mental Health, United States Public Health Service. The authors are indebted to Mrs. Deborah H. Nickerson for her assistance in this research.

[2] A portion of this paper was presented at the 1956 annual meetings of the Midwestern Psychological Association.

[3] Now at Adelphi College.

[4] An important aspect of procedure which was inadvertently omitted from the previous paper (**2**) was that all *S*s received *ten* daily periods of variable interval reinforcement.

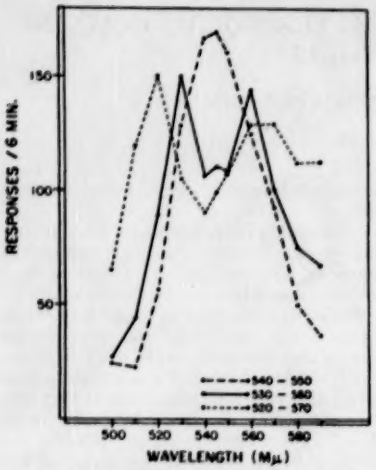


FIG. 1. Mean generalization gradients for the combined first and second tests.

RESULTS

The mean generalization gradients for the combined first and second tests are shown in Fig. 1. These were obtained by plotting the mean total number of responses for each test stimulus against wave length. The maximum rate of responding for the group trained on 540 and 550 Mμ occurs at 545 Mμ, while for the other two groups, the maxima occur in the regions of the training stimuli. It is apparent that the gradients obtained in the present experiment are basically consistent with those resulting from training to a single stimulus.

Further analysis of the present gradients, with respect to the question of summation, requires the formulation of exact alternative hypotheses based upon conceptions as to how gradients combine. Three methods of combination have been considered, and two of these will be presented in detail. At one extreme, it may be assumed that the total gradient for two stimuli consists of two independent underlying gradients which do not interact or summate such that the total gradient may be constructed merely by superpositioning two gradients of appropriate height at the CS values and tracing their outline. Figure 2 has been constructed in this manner. The method for obtaining the theoretical values in this figure will be described in detail later.

The second hypothesis to be con-

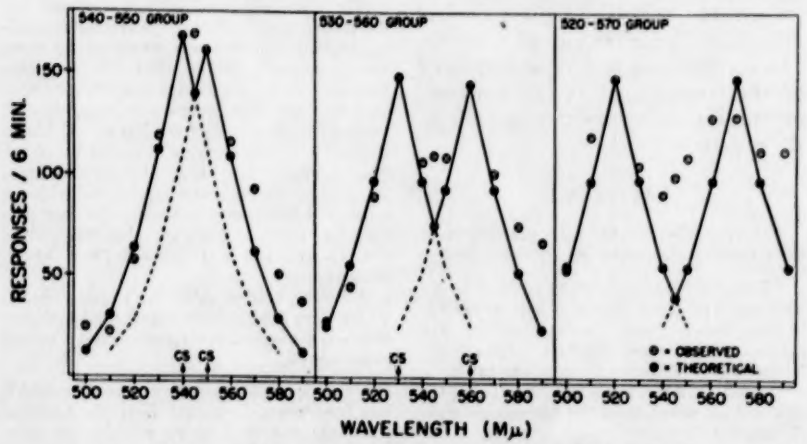


FIG. 2. Nonsummation hypothesis: theoretical curves and observed values.

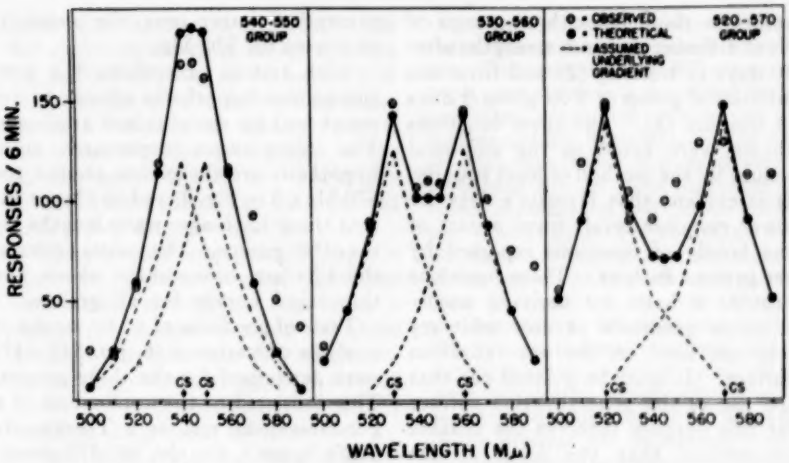


FIG. 3. Summation hypothesis: theoretical curves and observed values.

sidered stems from Hull's proposal that reactive tendencies such as those represented by generalization gradients summate exponentially. Theoretical curves based upon this assumption are shown in Fig. 3.

The third assumption is that gradients combine in a simple algebraic manner. A cursory examination of the results, particularly for the 540–550 group indicates that this hypothesis is untenable and excludes it from further consideration.

Derivation of hypothetical control gradients for single stimuli.—Analysis in terms of the previously discussed hypotheses requires a comparison with gradients obtained after training to a single stimulus. Since the present experiment provided no control data for animals trained to respond to a single stimulus, hypothetical generalization gradients of the appropriate parameters have been obtained from a "generalization surface" found in an earlier study (**2**) and given additional confirmation in an investigation by Hanson (**3**). It has been shown that there is a systematic change in the form of the generalization gradient for wave length as a function of response strength. The "generalization surface" (**2**; vide Fig. 5 and 6) describes these changes and this surface may be represented in a two-dimensional form as in Fig. 4 simply by connecting the values for a given test stimulus for gradients at various response strengths after these have been positioned in terms of area.

The empirical values in Fig. 4

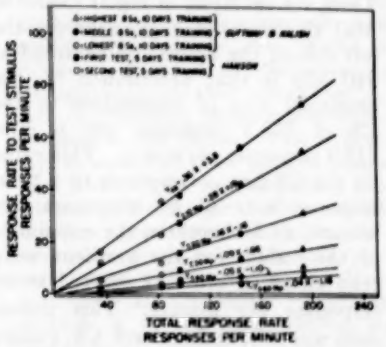


FIG. 4. Contours of the generalization surface for various test stimuli.

comprise those from three groups of *Ss* of different response strengths after 10 days of training (**2**) and from one additional group of 8 *Ss* given 5 days of training (**3**). The linear functions shown were fitted to the empirical values by the method of least squares. It is evident that Hanson's data for lower response levels form a part of the family of functions reported by the present authors. These functions provide a basis for deriving single-stimulus gradients at any arbitrary response level on the generalization surface. It must be pointed out that the use of the generalization surface for this purpose involves the tenable assumption that the form of the generalization gradient is independent of the spectrum locus of the CS (**2**). Additional justification for this assumption is provided by the present experiment in which gradients found around CS values, other than those used to generate the surface, are adequately represented by certain planes in the surface.

The nonsummation hypothesis.—The theoretical curves (Fig. 2) which constitute the predictions from the nonsummation hypothesis were obtained by selecting gradients at the response levels of the various CSs. Thus, for example, in Fig. 2 (540–550 Mμ) the underlying gradient for the left side of the figure and centered at 540 Mμ is that determined by the empirical rate of responding to the CS of 167.5 responses per 6 min. (27.9 responses per min.). This value on the CS line corresponds to a total response rate of 66 responses per minute, and, therefore, the remainder of the values for the gradient were read along an ordinate erected at 66 responses per minute. This procedure was followed for each CS, except for the gradient around 570 Mμ which is too irregular to permit an estimate of CS level. For this gradient the theoretical curve was the same as that used for 520 Mμ.

With certain exceptions, the nonsummation hypotheses appear to account well for the obtained gradients. The conspicuous departures from hypothesis are the points around the 570 Mμ CS in Fig. 2, and to a lesser extent those for longer wave lengths for the other groups. The values between the CSs are consistently above the theoretical curves for all groups.

Tests of goodness of fit by means of analysis of variance (**6**, pp. 344–347) were performed for the three groups. This consisted of the application of a goodness-of-fit test in a Treatments $\times$ *Ss* design. For the 540–550 group, the departure from hypothesis cannot be regarded as significant since $F = 1.20$ (8,60 *df*; $P > .20$). For the 530–560 group, a significant departure is obtained ($F = 2.45$; 8,60 *df*; $P = .025$). However, if the rate of responding for 590 Mμ is reduced to approximately that for 500 Mμ, the nonsummation hypothesis cannot be rejected ($F = 1.77$; 8,60 *df*; $P = .10$). For the 520–570 group, the nonsummation hypothesis can clearly be rejected ($F = 3.87$; 8,60 *df*; $P = .001$).

The Hullian summation hypothesis.—In order to apply Hull's theory to the present data, the problem is to determine whether the obtained gradients can be accounted for by exponential addition of two smaller gradients of equal magnitude taken from the generalization surface. The restriction of equal magnitude is required because equal training was given to both CSs, and a further restriction that the separate gradients fall on the generalization surface is imposed by the fact that this surface yields the best available prediction of the results for single-stimulus training.

In general, the underlying gradients chosen for summation (Fig. 3) were determined by a trial-and-error

method of least squares. Hull's rule for approximate exponential addition (**4**, p. 200) requires an estimate of M (the physiological maximum of habit strength attainable under optimal conditions). This was taken to be on the order of 180 key pecks per minute, and rate of key-pecking was interpreted directly as sH_R.

Upon inspection at least, it appears that the functions derived from the summation hypothesis constitute a fairly reasonable account of these results. In fact, the differences between the theoretical curves generated by the two hypotheses are surprisingly small in the present instance.

The same statistical tests of goodness of fit used with the nonsummation hypothesis were applied to the summation functions. The data for the 520–570 group depart significantly from the summation function ($F = 4.73$; 8,60 df; $P < .001$), just as in the case of the nonsummation function, and presumably for the same reasons. The summation function for the 530–560 group also departs significantly from the data ($F = 2.13$; 8,60 df; $.025 < P < .05$), but here again a downward adjustment of a single value (590 Mμ) serves to reduce the F to 1.13 (8,60 df; $P > .20$). In the case of the 540–550 group, the summation hypothesis must be rejected ($F = 2.18$; 8,60 df; $.025 < P < .05$) in contrast to the finding for the nonsummation hypothesis. However, as before, the rejection of the hypothesis of goodness of fit hinges upon a single value since a reduction in the rate of responding for 590 Mμ results in a F of 1.95 (8,60 df; $.05 < P < .10$).

Discussion

While this experiment has shown that the generalization gradient for two stimuli is some combination of two single-stimulus gradients, the results can be accounted for nearly equally well by the summation and nonsummation hypotheses. The quantitative differences between the theoretical functions for the conditions of this experiment are extremely small. Moreover, on the basis of inspection, there is a good over-all correspondence in most instances between experimental and theoretical values. Although in five out of six cases the statistical tests require the rejection of the hypothesis of good fit, in three of these instances (for the 540–550 group the summation function, and for the 530–560 group both functions) the rejection may be regarded as contingent upon a single peripheral value. It would seem unreasonable to suggest a choice between the hypotheses without more substantial evidence.

There are nonetheless certain details in the findings of the present experiment which, taken in conjunction with the results of Bilodeau, Brown, and Meryman, tend to influence one's speculations as to how generalization gradients combine. In the present study, as in the Bilodeau, Brown, and Meryman experiment, the only obvious suggestion of summation appears to occur in the region between the training stimuli. For all three groups in the present study, the center values only are consistently raised above hypothetical gradients erected at the exact values of the CSs. Despite the fact that two of these three differences (540–550, 530–560) are not statistically significant, the pattern nevertheless conforms to that found by Bilodeau, Brown, and Meryman and strongly suggests that summation of a sort is a genuine phenomenon and is confined to this region. This is not to say, however, as these authors suggest, that what occurs in this region is an instance of Hullian summation. A sufficient and precise test of Hull's rule of summation must take into account *all* overlapping values of the combining gradients. If Hull's rule were rigorously applied to the Bilodeau, Brown, and Meryman experiment, it would be seen that in five of the six instances where control gradients overlap and summation might be expected, none occurs. In

view of this, the interpretation of the Bilodeau, Brown, and Meryman experiment as providing support for the Hullian theory of summation seems unwarranted.

The present results suggest the experimental paradigm which would be likely to furnish a definite answer to the problem of how generalization gradients combine. An appropriate procedure would consist of training to three stimuli spaced approximately ten millimicrons apart in order to provide a sufficient degree of double overlap. In this situation it can be expected that predictions based upon the alternative hypotheses would be sufficiently different so that they could be distinguished by the present technique.

The form of the generalization surface.—In a previous paper (2), the authors had suggested that the generalization surface could be adequately described by a function with both additive and multiplicative components. The inclusion of Hanson's (3) results and the fitting of linear equations (Fig. 4) to the combined sets of points clearly necessitates a revision of the earlier hypothesis. At the present time, the findings are almost completely compatible with a pure multiplicative relationship, since the intercept constants of all six equations are very close to zero. The greatest departure from zero (+3.8 responses per min.) occurs in the case of the equation for the CS contour. It is quite possible that this intercept constant is too large and its associated-slope constant too small due to a systematic underestimation of the peak gradient values. Evidence to be presented in a later study indicates that, in general, the peak values cannot be assumed to be located at the CS.

Summary

An experiment was performed to explore the problem of the summation of stimulus-generalization gradients. Three groups of pigeons were reinforced for key-pecking at a disk illuminated by monochromatic light. One group was given an equal number of reinforcements with wave lengths of 540 and 550 Mμ; a second group was trained with 530 and 560 Mμ, and a third group with 520 and 570 Mμ. Generalization gradients for all groups were obtained over the range from 500 to 590 Mμ from measures of response rate during extinction. The gradients around each CS generally resembled those obtained after training to a single stimulus.

The results were compared with alternative hypotheses (Hullian exponential summation vs. nonsummation) regarding the manner in which overlapping gradients combine. Use was made of a set of empirical functions relating the form of the single-stimulus generalization gradient to response strength ("generalization surface") to derive the two-stimulus gradients required by the hypotheses. The two theories of combination appeared to account for the obtained results almost equally well. An extension of the method of the present study to afford a more conclusive answer to the problem was suggested.

With additional data, the generalization surface was found to be more nearly in accord with the notion that changes in the generalization gradient with changes in response strength can be described by means of a simple multiplicative relation.

References

1. Bilodeau, E. A., Brown, J. S., & Meryman, J. J. The summation of generalized reactive tendencies. *J. exp. Psychol.*, 1956, **51**, 293–298.
2. Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.
3. Hanson, H. M. The effects of discrimination training on stimulus generalization. Unpublished doctor's dissertation, Duke Univer., 1956.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
5. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1936, **46**, 9–30.
6. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
7. Wolfle, D. Training. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Chap. 34.

(Received for early publication August 17, 1956)

Journal of Experimental Psychology
Vol. 53, No. 2, 1957

DYNAMIC CONTOUR PERCEPTION [1]

W. M. SMITH AND W. L. GULICK

Princeton University

Contour in visual perception has not received an amount of attention deserved by the fundamental role it plays in all visual activities. One investigator has indicated his view of the importance of contour by writing, " . . . the matter of contour formation is one of the central phenomena of sensory psychology and nerve physiology" (**1**, p. 334). If visual acuity, form perception, visual contrast, and stereopsis depend primarily upon the ability of the visual system to "construct" an edge from varying patterns of energy that strike retinal receptors, then it is clear that contour must be systematically investigated if we seek to understand any visual activity.

Contour is usually defined as an abrupt gradient of stimulation of adjacent areas on the retina, and this gradient is dependent upon intensity differentials in the objective field (**2, 11**). Ordinarily this definition is adequate but phenomena of "subjective contour," certain contrast effects, and the perception of edges of *apparently* moving stimuli are instances which require an extension of the definition. Perceived contour does not always depend upon an abrupt gradient in stimulation even though this is the usual stimulus correlate (**5, 6**).

Time is required to form visual contours. The exact magnitudes for various conditions of observation, however, have not been studied systematically. Werner, nevertheless, has shown that the contours of certain geometric figures presented stroboscopically for 12–25 msec. are obliterated when an enclosing figure with a common center follows in 120–240 msec. (**13**). In most instances the first figure was not seen at all. This study, unfortunately, does not give systematic data on the appearance of contour of the first figure as a function of its duration of exposure. It must be assumed that a duration greater than 25 msec. would produce contour. A second study relating to time factors in the perception of contour is one by Cheatham (**2**). He indicated that the development time of perception (the perception of complete contours) as measured with a masking method is of the order of 30–100 msec., depending upon the shape of the stimulus and its brightness.

In the present experiment the time required to perceive the contours of moving stimuli was studied. Heretofore, this topic had not been investigated. For the purposes of the present study, a distinction is made between static and dynamic contour perception, the latter referring to the formation and maintenance of sharp edges during the actual movement of a stimulus. This distinction parallels the one made by Miller and Ludvigh between static and dynamic visual acuity (**9**).

Method

First it is necessary to describe the phenomenon upon which observations in this study are based. A moving stimulus is presented at some velocity (V) which makes it impossible to perceive sharp contours. Typically, contours

[1] This research was supported by a grant from the National Science Foundation to the senior author.

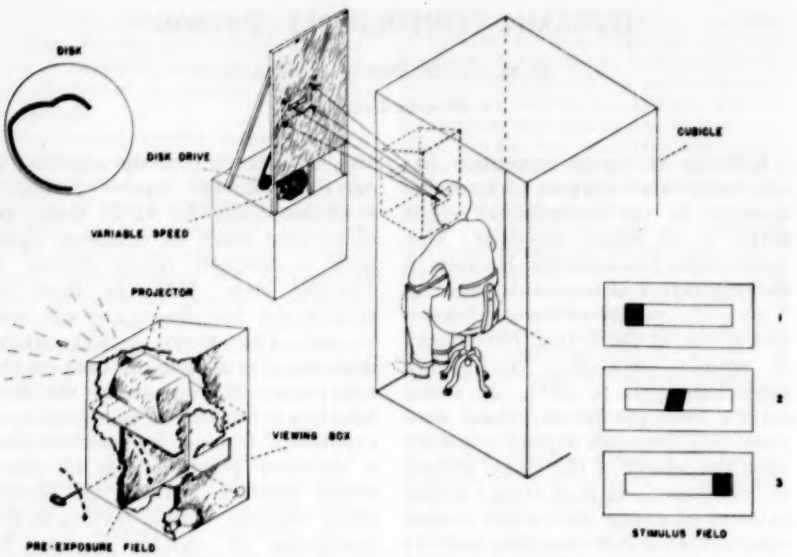


FIG. 1. Apparatus components and features of the observed stimulus.

of small stimuli (a few degrees in angular size) no longer can be discerned when V exceeds 15°/sec. (4). Now suppose that $V = 20°$/sec. and the stimulus is exposed in a fixed position for a brief interval prior and subsequent to its movement. Under these conditions the previously blurred edges of the stimulus now are seen in sharp contour, as they were when $V < 15°$/sec.[3]

With respect to contour perception, the present study investigated the relation between values of V and durations of exposure of the stimulus in its fixed position before and after movement. Henceforth t_1 and t_2 will designate the duration of the exposure of the stimulus in its fixed positions prior and subsequent to movement, respectively. The letter T will indicate the duration of exposure when the relationship $t_1 = t_2$ holds.

Apparatus.—A device similar to the one employed by Michotte was used (8). By rotating a large disk on a horizontal axis behind an aperture, numerous combinations of t_1 and t_2 could be obtained. The velocity of the stimulus was altered by changing the speed of the disk's rotation. A variable speed transmission (Graham) was used to drive the disk. Exposure of t_1 and t_2 was controlled both by the speed of the disk and the length of the concentric arcs upon it. The schematic drawings in Fig. 1 illustrate the important components of the apparatus and the general features of the experimental situation. A drawing of the disk is given in the upper-left portion of the figure. Rotation of the disk behind a horizontal aperture produced a stimulus field which appeared like the set of three drawings in the bottom-right part of the figure. Drawing 1 illustrates the field when the stimulus first appeared; Drawing 3 shows the field after movement had taken place. The slanted band in Drawing 2 shows the appearance of the stimulus during movement. The slant is due to the "fall" of the eccentric arc on the disk.

The stimulus field was observed by S from a cubicle. He looked through a viewing box. The upper half of the box contained a projector which illuminated the stimulus field. In the lower half of the viewing box a white movable panel served as the pre-exposure field. The S always looked at the pre-exposure field between observations.

The size of the white stimulus field was 8°18′ by 1°30′. Its brightness was 50 ft.-l. The larger area surrounding this field and the concentric and eccentric bands on the disk were all

[3] Credit for this original observation is due to Professor Albert Michotte of the Psychological Laboratory of Louvain University, Louvain, Belgium.

painted black, having a brightness of 1.5 ft.-l. All measures of brightness were made with a MacBeth Illuminometer. Movement extended over 5° of visual angle and the stimulus subtended an angle of approximately 0°30′.

Instructions.—The instructions for observing the moving stimulus when $T = 0$ directed S to report contour if the stimulus appeared with sharp edges and homogeneous blackness throughout the extent of movement. Neither a fixation point nor instructions regarding visual fixation and pursuit were given S. The criterion of sharpness was demonstrated by showing S the stimulus both in a fixed position and as it moved across the field at a low velocity. This sequence also was shown periodically to remind S of the criteria to be met. For those trials in which t_1 and $t_2 > 0$, S was given similar instructions regarding contour. On any trial where the appearance of the stimulus did not meet the desired criteria, S was requested to describe what he saw. Preliminary data indicated that in spite of the difficult and complex task set for S, each was able to maintain the criteria of sharp edges and homogeneous blackness with remarkable stability.

Procedure.—During 5 min. of light adaptation to the pre-exposure field E read a standard set of instructions to S.[3] Then the demonstrations regarding the criteria of contour were given.

The stimulus was shown to S at a low velocity with $T = 0$. Four successive presentations of the moving stimulus were given, after which S stated the presence or absence of contour. By following a modified method of limits, V was increased until S reported the absence of contour. At this V positive values of t_1 and t_2 were introduced in 10–20 msec. steps. The magnitude of t_1 and t_2 was increased progressively until S reported the reappearance of contour. The next step involved increasing V in 1–2° steps until contour was not observed. Then, once again, the exposure duration of t_1 and t_2 was increased until contour was regained. In this manner a series of values was obtained for each S which represented the duration of T and enabled S to report the presence of contour during movement of the stimulus for different velocities. Data for both horizontal and vertical movement were gathered, the latter type of movement being presented simply by modifying the orientation of the aperture in front of the rotating disk.

In order to determine the differential effects of t_1 and t_2 on the perception of contour, data also were obtained on trials over the range of values for V where $t_1 \neq t_2$. The procedure used was as follows: At combinations of V and T which were adequate to produce contour, t_1 (or t_2) was reduced in duration to .00, .25T, .50T, and .75T in that order, while t_2 (or t_1) remained equal to T. The S made a judgment of contour at each of these combinations of t_1 and t_2. For example, assume that when $V = 17$°/sec. S required a T of 200 msec. before reporting contour. With $t_2 = 200$ msec., t_1 was reduced to 0, 50, 100, and 150 msec. Then with $t_1 = 200$ msec., t_2 was reduced to 0, 50, 100, and 150 msec. For each of these eight combinations S made a judgment of contour. On odd-numbered trials t_2 was altered after t_1, whereas on even-numbered trials the reverse was true.

Data from five graduate students are included in the results. All Ss were considered "sophisticated" because each had made numerous preliminary observations prior to the systematic ones reported here. Each S had 20/20 vision, either corrected or uncorrected.

Results

The data of each of the five Ss for horizontal movement are given in the individual graphs in Fig. 2. In all instances $t_1 = t_2$. The time exposure of the stimulus in a fixed position before and after movement is plotted against velocity. A zero value of T means that the stimulus was moving as it appeared and disappeared. Because of the procedure employed, not all of the functions are defined by the same number of points. Visually fitting a straight line to the data for each S clearly indicates that within the range of values of V and T studied, the contour of the moving stimulus can be maintained when V increases as long as T increases. The apparatus did not permit exposures of t_1 and t_2 greater than 350 msec. Consequently, nothing can be said about the nature of the relation between higher velocities of movement and increased magnitudes of T. Obviously, however, at and beyond some critical velocity contour will not be seen, regardless of the duration of T.

[3] The brightness of the pre-exposure field was planned to be equal to that of the white part of the stimulus field. Due to an error in notation its brightness was approximately 100 ft.-l.

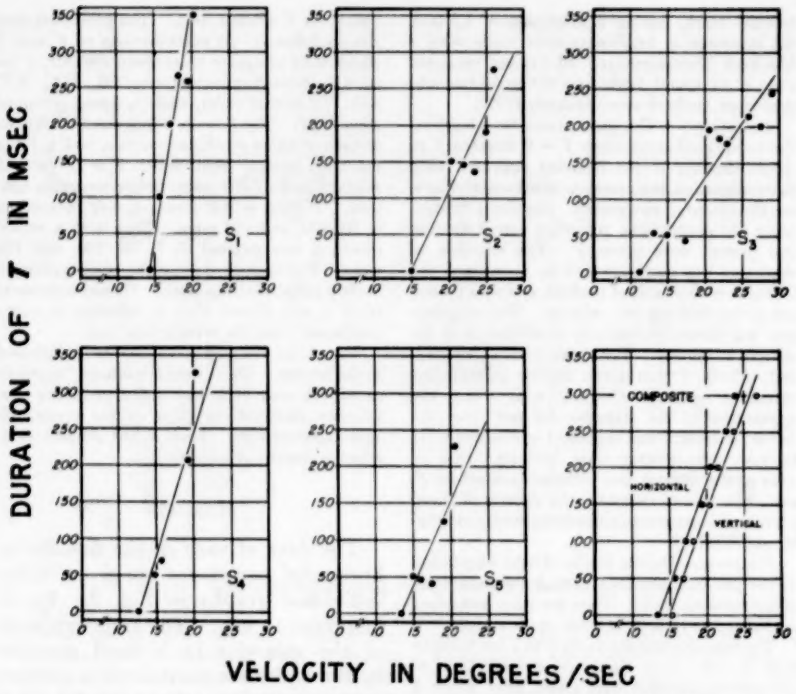


FIG. 2. Individual and composite *contour-contours*. The individual functions represent data for horizontal movement only.

The individual functions in Fig. 2 may be described as *contour-contours*. All the combinations of T and V to the left of the lines produce the perception of sharp edges. Those combinations to the right do not. Considering the difficulty of maintaining the criteria of judgment, the data for each of the five *S*s are remarkably similar. In some individual graphs there are clusters of two and three points. These data reflect reversals of judgment which occurred when V was increased after a given combination of V and T had produced contour. This increase in V had the effect of reducing the value of T unless compensated for by an extension of the concentric arcs on the disk. Reversals in the data mean, therefore, that contour reported for a particular combination of V and T was also reported when V was increased and T was decreased slightly.

A composite of individual functions for horizontal and vertical movement is shown in the lower-right graph in Fig. 2. These composite functions were derived by interpolating values of velocity from the five best-fitting lines at 50-msec. intervals of T, beginning with $T = 0$. The means for the five interpolated values of V for each 50-msec. interval of T represent the composite functions. The perception of contour for stimuli moving horizontally and vertically is very similar. The small but consistently

higher values of V for vertical movement probably result from a practice effect since these observations were made after those of horizontal movement. The two functions show that visual contour can be maintained if for every increase of 1°/sec. in V the exposure of T is increased by approximately 30 msec.

The effects on contour when $t_1 \neq t_2$ are shown in Fig. 3. Data for both horizontal and vertical movement are presented. In this figure the percentage of trials in which contour was reported has been plotted against the relative durations of t_1 and t_2. Data for all *S*s and all velocities of movement have been combined. The reference value of $^1/_1$ is based upon the number of trials where $t_1 = t_2$ and contour was perceived. Accordingly, for horizontal movement contour was reported on approximately 80% of the trials in which t_1 equalled some positive value and $t_2 = .50$ of t_1. The

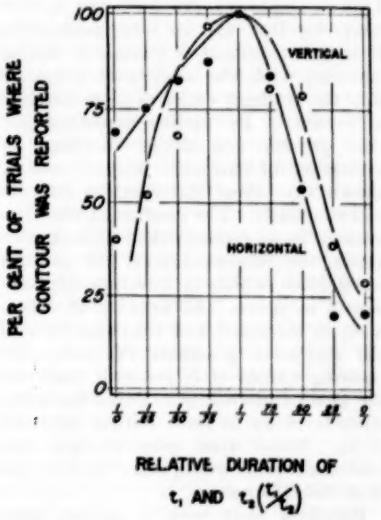


FIG. 3. Curves showing the relation between contour perception and the relative values of t_1 and t_2.

asymmetry of the curves, at least of the one for horizontal movement, indicates that t_1 was relatively more important for the perception of contour than was t_2. While the figure does not show it, a separate study of the effects of t_1 and t_2 for different specific velocities of movement suggests that the importance of t_1 over t_2 is more pronounced at the higher velocities. Figure 3 does indicate, however, that in the absence of either t_1 or t_2, but not both, contour is still observed a certain per cent of the time. This means that with an exposure of the stimulus in a fixed position either before or after movement, contour can be seen at velocities greater than in those cases where only a moving stimulus is presented. Data of individual *S*s indicate clearly that both t_1 and t_2 are necessary for contour perception at the very high velocities.

DISCUSSION

The data show that dynamic contour perception is influenced to an important degree by the events preceding and following stimulus movement. The edges of a stimulus which are not seen clearly at a particular velocity can be observed if the stimulus is presented in a stationary position before and after its movement. The explanation of this phenomenon is not obvious on the basis of present knowledge of the visual system and its functions.

Because of the sequence of stimulation in the procedure employed in the present experiment, the results might be explained on the basis of *optimal* apparent movement. By omitting the second stage of the sequence, i.e., real movement, the usual conditions for presenting apparent movement are achieved. It is possible, therefore, that *S*s were reporting the contours of an *apparently* moving stimulus rather than those of one in actual motion. There are, however, two principle arguments against this interpretation. One stems from the fact

that the stimulus during movement was tilted. Inability to observe this particular orientation would be presumptive evidence that *S* was perceiving *optimal* apparent movement rather than real movement. But this apparently was not the case here since the tilt of the stimulus during movement could be observed.[4] The second point of argument is based on the obvious difference in appearance of movement and contour with and without the physically moving stimulus. Comparative observations of this difference were not systematically made, but observations by some *S*s indicated that the difference was obvious. By using a second disk which allowed the presentation of t_1 and t_2 without the moving phase, several combinations of durations of t_1 and t_2, and the interval between them produced various degrees of apparent movement. This apparent movement included "phi" and was always less than optimal. Less than optimal movement would be expected when one considers the relatively long intervals which separated t_1 and t_2.[5] Although the results obtained in this study are probably independent of *optimal* apparent movement, the greater importance of t_1 as compared to t_2, at least for horizontal movement, certainly suggests the relevance of apparent movement to these data (**10**).

Data on the effects of the relative duration of t_1 and t_2 show that visual contour at increasing velocities can be maintained to some extent when either t_1 or $t_2 = 0$, the former being much more important for contour resolution than the latter. Obviously, apparent movement is not involved under these conditions.

The differential importance of t_1 and t_2, the relationship between duration of T and magnitude of V, plus the temporal and spatial features of the stimulus presentations suggest a tentative explanation of the present results in terms of a statistical conception of contour resolution. This conception can be traced to the work of Marshall and Talbot (**7**), and recently expanded by Osgood and Heyer (**12**) to account for figural aftereffects. More recently, Day (**3**) has used the essential ideas of statistical theory to explain form perception. One basic feature of this theory is that a contour or edge is represented in the visual cortex as a " . . . normal distribution of excitation, symmetrical about its axis transversely and extending as a ridge throughout the longitudinal extent of the contour" (**12**, p. 102). Parameters of this hypothetical distribution are considered to be affected by the state of adaptation, contrast, and the intensity of illumination.

It may be assumed here that the excitation gradient representing a visual contour must reach or surpass some minimum level before the edge of a stimulus, fixed or moving, can be seen. This assumption is the same one as that made by Day (**3**) for form perception. It is here extended to include a moving contour, with the additional provision that the gradient varies as some function of velocity. By further assuming specific growth and decay functions of excitation for movement and t_1 it can be shown that these distributions interact and summate. The product of the summation is a single-peaked distribution above the minimum level for contour recognition which is continuously displaced in space. As velocity of movement is increased and thus the level of the excitation gradients decreased, increasing values of t_1 can still maintain the peaked distribution above the minimum level up to some critical duration of t_1. Based upon present data this critical duration would seem to be no less than 300–350 msec.

Recalling that both t_1 and t_2 were necessary for contour resolution at the higher velocities, it is also necessary to indicate the role played by t_2 in the

[4] One *S* at a velocity of 30°/sec. reported the absence of slant of the stimulus, indicating optimal apparent movement was being observed.

[5] A transmission setting corresponding to V of 30°/sec. represented a time interval between t_1 and t_2 of approximately 166 msec. At 20°/sec. the interval was about 250 msec., and for 14°/sec. it was about 357 msec.

process described above. By using the statistical basis of contour for *apparent movement*, as suggested by Osgood (11, pp. 243–248), it may be assumed that overlapping distributions in the cortex, representing stimuli separated in space, can produce under proper conditions a single-peaked distribution continuously shifting in cortical space. The perceived effect of this state of affairs is optimal apparent movement. Since the higher velocities studied in the present experiment were of such magnitudes that the time interval between t_1 and t_2 was 200 msec. or slightly less, the conditions for apparent movement were being approached if not achieved. It is concluded, therefore, that the necessary presence of t_2 for contour perception at the higher velocities represents the combined effects of the excitation distribution for real movement and for apparent movement. At the lower velocities the summed interaction of t_1 and real movement maintained contour at velocities greater than that possible without t_1. At the higher velocities the summation was between the contour processes of real and apparent movement. It is evident that at velocities slightly above 30°/sec. the excitation gradient from real movement would be negligible, while that for apparent movement would be maximal. In this instance the perceived contour of the stimulus moving from t_1 to t_2 would be apparent. At extremely high velocities neither real nor apparent contours would be discernible; instead, only simultaneity of two discrete stimuli separated in space.

It is obvious that the above discussion neglects many important details of neurophysiological activity which must underlie the perceptual results described as well as complicating factors of visual fixation, tracking movements of the eyes, and angular extent of movement. However, the results and their interpretation presented here are compatible with the statistical theory of perceptual processes in its present formulation. The details of the quantitative relationships indicated above currently are being worked out, and data concerning some of the complicating factors mentioned are being gathered.

Summary

In this experiment the perception of contour of a moving stimulus was studied in relation to how long the stimulus was seen before or after it moved, or both. On the basis of observations from five *S*s involving velocities of movement up to 30°/sec., the following conclusions and interpretations were made:

1. The sharp contour of a moving stimulus cannot be observed when its velocity exceeds a value of about 10–15°/sec.
2. By exposing a stimulus in a fixed position before and after movement, contour can be maintained during movement at least up to velocities of 25–30°/sec.
3. For the production of contour during movement, the relation between the duration of exposure of the stimulus in fixed positions before and after movement and the velocity of the stimulus is linear, within the limits studied.
4. Contour is maintained most effectively under the conditions studied when movement is both preceded and followed by equal magnitudes of exposure of the stimulus in fixed positions.
5. A general explanation of the results was proposed on the basis of the statistical theory of perceptual processes.

References

1. Bartley, S. H. *Vision: A study of its basis.* New York: Van Nostrand, 1941.
2. Cheatham, P. G. Visual perceptual latency as a function of stimulus brightness and contour shape. *J. exp. Psychol.*, 1952, **43**, 369–380.
3. Day, R. H. Application of the statistical theory to form perception. *Psychol. Rev.*, 1956, **63**, 139–148.
4. De Silva, H. R. An analysis of the visual perception of movement, *Brit. J. Psychol.*, 1929, **19**, 268–305.
5. Ludvigh, E. J. Perception of contour: I. Introduction. U. S. Naval School of Aviation Medicine, Joint Proj. Rep. No. 4, NM 001 075.01.04, 1953, 11 pp.
6. Ludvigh, E. J. Perception of contour: II. Effect of rate of change of retinal intensity gradient. U. S. Naval School of Aviation Medicine, Joint Proj. Rep. No. 5, NM 001 075.01.05, 1953, 9 pp.
7. Marshall, W. H., & Talbot, S. A. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. In H. Kluver (Ed.), Visual

Mechanisms, Biol. Sympos., 1942, **7**, 117–164.

8. MICHOTTE, A. La perception de la causalité. Louvain: Institut Supérieur de Philosophie, 1946.
9. MILLER, J. W., & LUDVIGH, E. J. Dynamic visual acuity when the required pursuit movement is in a vertical plane. Report No. 2, USN Bureau of Medicine and Surgery, Washington, D. C., 1953.
10. NEUHAUS, W. Experimentelle Untersuchung der Scheinbewegung. *Arch. ges. Psychol.*, 1930, **75**, 315–458.
11. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
12. OSGOOD, C. E., & HEYER, A. W., JR. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.
13. WERNER, H. Studies of contour: I. Qualitative analysis. *Amer. J. Psychol.*, 1935, **47**, 40–64.

(Received for early publication August 17, 1956)

ERRATUM

In the article "Further Tests of Theories of Decision in an 'Expanded Judgment' Situation" by Francis W. Irwin and W. A. S. Smith in the December 1956 issue of this JOURNAL (*J. exp. Psychol.*, 1956, **52**, 345–348), there are two errors of wording in the text. In line 7 of RESULTS AND DISCUSSION (p. 347) and again in line 7 of the SUMMARY (p. 348), the word "directly" should be changed to "inversely." The word "inversely" was intended in both instances, and was required by the results and by the two theories under discussion.